



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA PODNIKATELSKÁ
ÚSTAV MANAGEMENTU**

FACULTY OF BUSINESS AND MANAGEMENT
INSTITUT OF MANAGEMENT

VYUŽITÍ PROSTŘEDKŮ UMĚLÉ INTELIGENCE NA KAPITÁLOVÝCH TRZÍCH

THE USE OF MEANS OF ARTIFICIAL INTELLIGENCE FOR DECISION MAKING SUPPORT
ON STOCK MARKET

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. VLASTIMIL HRACH

VEDOUCÍ PRÁCE
SUPERVISOR

prof. Ing. PETR DOSTÁL, CSc.

BRNO 2011

ZADÁNÍ DIPLOMOVÉ PRÁCE

Hrach Vlastimil, Bc.

Řízení a ekonomika podniku (6208T097)

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách, Studijním a zkušebním řádem VUT v Brně a Směrnicí děkana pro realizaci bakalářských a magisterských studijních programů zadává diplomovou práci s názvem:

Využití prostředků umělé inteligence na kapitálových trzích

v anglickém jazyce:

The Use of Means of Artificial Intelligence for the Decision Making Support on Stock Market

Pokyny pro vypracování:

Úvod

Vymezení problému a cíle práce

Teoretická východiska práce

Analýza problému a současné situace

Vlastní návrhy řešení, přínos návrhů řešení

Závěr

Seznam použité literatury

Přílohy

Seznam odborné literatury:

- ALIEV, A., ALIEV, R. Soft Computing and Its Applications. World Scientific Pub. Ltd, 2002. 444 s. ISBN 981-02-4700-1.
- DOSTÁL, P. Pokročilé metody analýz a modelování v podnikatelství a veřejné správě. 1. vyd. Brno : CERM, 2008. 340 s. ISBN 978-80-7204-605-8.
- DOSTÁL, P., SOJKA, Z. Financial Risk Management. Zlín : UTB, 2008. 80 s. ISBN 978-80-7318-772-9.
- DOSTÁL, P. Advanced Economic Analyses. Brno : CERM, 2008. 80 s. ISBN 978-80-214-3564-3.
- HANSELMAN, D., LITTLEFIELD, B. Mastering MATLAB7. Pearson Education International Ltd., 2005. 852 s. ISBN 0-13-185714-2.
- REJNUŠ, O. Finanční trhy. Ostrava : Key Publishing, 2008. 559 s. ISBN 978-80-87071-87-8.
- THE MATHWORKS. MATLAB – Neural Network Toolbox - User's Guide. The MathWorks Inc., 2010.

Vedoucí diplomové práce: prof. Ing. Petr Dostál, CSc.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2010/2011.

L.S.

PhDr. Martina Rašticová, Ph.D.
Ředitel ústavu

doc. RNDr. Anna Putnová, Ph.D., MBA
Děkan fakulty

V Brně, dne 23.05.2011

Anotace

Tato diplomová práce popisuje využití umělé inteligence pro predikci na akciových trzích. Predikce staví netradičně na pravděpodobnostním modelu Bayesova vzorce a na něm založeném naivním Bayesově klasifikátoru. V praktické části je proto navržen algoritmus, který pro odhad budoucího vývoje akcie používá rozpoznané vztahy mezi identifikátory technické analýzy. Konkrétně se jedná o exponenciální klouzavé průměry za 20 a 50 dní. Na základě klasifikace vztahů mezi identifikátory je výstupem programu grafický odhad budoucího vývoje akcie.

Annotation

The diploma thesis deals with artificial intelligence utilization for predictions on stock markets. The prediction is unconventionally based on Bayes' probabilistic model theorem and on its based Naïve Bayes classifier. In the practical part algorithm is designed. The algorithm uses recognized relations between identifiers of technical analyze. Concretely exponential running averages at 20 and 50 days had been used. The program output is a graphic forecast of future stock development which is designed on ground of relations classification between the identifiers

Klíčová slova

Akcie, predikce, umělá inteligence, strojové učení, pravděpodobnost, hypotéza, Bayes, naivní bayesův klasifikátor, klouzavý průměr

Key words

Stock, prediction, artificial intelligence, machine learning, probability, hypothesis, Bayes, naive bayesian classifier, moving average

Bibliografická citace práce:

HRACH, V. *Využití prostředků umělé inteligence na kapitálových trzích*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2011. 88 s. Vedoucí diplomové práce prof. Ing. Petr Dostál, CSc.

Čestné prohlášení

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/200 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 23. 5. 2011

Poděkování

Děkuji vedoucímu diplomové práce prof. Ing. Petru Dostálovi, CSc. za odbornou a metodickou pomoc při hledání řešení problémů zpracování mé diplomové práce. Dále děkuji recenzentovi Ing. Janu Budíkovi za mnoho užitečných rad.

Obsah

| | |
|---|----|
| Úvod a cíle práce | 10 |
| 1 Teoretická východiska | 11 |
| 1.1 Finanční trh, finanční systém | 11 |
| 1.2 Rozdělení finančních trhů | 12 |
| 1.2.1 Peněžní trh | 12 |
| 1.2.2 Kapitálový trh | 13 |
| 1.3 Predikce a akciové trhy | 15 |
| 1.4 Analýza akcie | 16 |
| 1.4.1 Fundamentální analýza | 16 |
| 1.4.2 Psychologická analýza | 20 |
| 1.4.3 Teorie spekulativních bublin | 21 |
| 1.4.4 Technická analýza | 21 |
| 1.4.5 Technické indikátory | 22 |
| 1.5 Umělá inteligence a strojové učení | 24 |
| 1.5.1 Zařazení strojového učení | 24 |
| 1.5.2 Umělá inteligence | 25 |
| 1.5.3 Strojové učení | 27 |
| 1.5.4 Statistika | 27 |
| 1.5.5 Optimalizace | 27 |
| 1.6 Komponenty procesu modelování | 28 |
| 1.6.1 Základní schéma | 28 |
| 1.6.2 Veličiny | 29 |
| 1.6.3 Modely | 30 |
| 1.7 Chybové funkce | 30 |
| 1.7.1 Metoda nejmenších čtverců | 30 |
| 1.7.2 Lineární MSE | 31 |
| 1.7.3 Váhová MSE | 32 |
| 1.7.4 Přeučení modelu | 32 |
| 1.8 Thomas Bayes (1702-1761) | 33 |
| 1.8.1 Bayesův vzorec | 33 |
| 1.8.2 Základní formulace Bayesova vzorce | 35 |
| 1.9 Bayesův vzorec – základní vztahy | 36 |
| 1.10 Bayesovské učení | 37 |
| 1.11 Naivní Bayesův klasifikátor | 38 |
| 1.12 Naivní Bayesův pravděpodobnostní model | 39 |
| 2 Analýza problému a současná situace | 41 |
| 3 Vlastní návrh řešení – implementace NBK | 42 |
| 3.1 Programové prostředí MATLAB | 42 |
| 3.2 NBK - algoritmus | 43 |
| 3.2.1 Prerekvizity z teoretické části | 43 |
| 3.3 NBK implementace | 45 |
| 3.3.1 Data pro obecnou implementaci | 45 |
| 3.3.2 NBK – vývojový diagram algoritmu | 47 |
| 3.3.3 NBK – popis algoritmu | 48 |
| 3.3.4 NBK – vlastní algoritmus | 48 |
| 3.4 Testování na obecných datech | 56 |

| | | |
|--------|--|----|
| 3.5 | TEST 1 – 3x5 | 57 |
| 3.5.1 | TEST 1.1 | 58 |
| 3.5.2 | TEST 1.2 | 60 |
| 3.5.3 | TEST 1.3 | 63 |
| 3.6 | TEST 2 – 3x9 | 65 |
| 3.6.1 | TEST 2.1 | 66 |
| 3.7 | Zhodnocení obecné části | 68 |
| 3.8 | NBK – použití pro reálná burzovní data | 68 |
| 3.8.1 | Kandidát pro predikci | 69 |
| 3.8.2 | Data | 69 |
| 3.8.3 | Normalizace dat | 71 |
| 3.9 | Testování na reálných datech | 73 |
| 3.10 | TEST 3 – 10x10 | 73 |
| 3.10.1 | TEST 3.1 | 74 |
| 3.11 | TEST 4 – 5x30 | 75 |
| 3.11.1 | TEST 4.1 a 4.2 | 75 |
| 3.11.2 | TEST 5 – 5x60 | 78 |
| 3.12 | Praktické použití pro investice | 80 |
| | Závěr | 84 |
| | Seznam použité literatury | 85 |
| | Seznam zkratk | 87 |
| | Seznam obrázků | 88 |

Úvod a cíle práce

Cílem práce je vytvoření algoritmu, využívajícího Naivní Bayesův klasifikátor, tedy pravděpodobnostní model, který dává do souladu pozdější zkušenosti s původními předpoklady. Tento model se pokusím aplikovat pro prediktivní odhad ceny akcií. Je tedy nutno se seznámit se základními pojmy ze strojového učení a umělé inteligence, s koncepcí procesu modelování, vysvětlit princip Bayesova vzorce a jeho aplikaci pro Bayesovský klasifikátor. Následná aplikace algoritmu na reálná data vyžaduje seznámení s metodami používanými při predikci akciových trhů a volbu vhodných ukazatelů pro modelové zpracování. V mém případě se jedná o indikátory technické analýzy, a to o exponenciální klouzavé průměry různé délky – typicky EMA 20, EMA 50 a EMA 200. Posláním této práce je dokázání vhodnosti použití modelu, který je chováním blízký lidskému uvažování.

1 Teoretická východiska

V této části práce nastíním základní strukturu finančních trhů, potažmo kapitálových trhů, kterých se týká praktická část diplomové práce. Jejich nejrozšířenější členění a identifikaci, metody analýz a indikátory. Je nutné se seznámit s interpretací výsledků těchto analýz. Pro vytvoření programu je potřeba se seznámit s principy strojového učení, potažmo umělé inteligence. Poslední částí teoretických východisek je pak seznámení se stěžejní částí, a to s popisem Bayesova vzorce a na něm postaveném Naivním Bayesově Klasifikátoru.

1.1 Finanční trh, finanční systém

Finanční trh je jedním z několika článků komplexu trhů, které rozlišujeme v rámci tržní ekonomiky. Samotný finanční trh je institucionální mechanismus, který byl vytvořen společnostmi za účelem směřování úspor a jiných finančních služeb k těm jednotlivcům a institucím, které jsou ochotny za ně platit. Finanční trh je centrem celého systému. Určující objem dosažitelných úvěrů, podporující spořivost, vytvářející úrokové sazby a hodnoty cenných papírů.(13)

Největším přínosem finančního trhu je přesun volných finančních prostředků z místa jejich vzniku do oblastí, kde je jejich potřeba. Jedná se o místo, kde dochází k přesunu volných finančních prostředků od přebytkových subjektů k subjektům deficitním a to především k těm, kteří pro ně mají nejefektivnější využití. Přebytkové subjekty mají v dané době nadbytek finančních prostředků, a proto mají vůli zapůjčit tyto volné zdroje. Deficitní subjekty se pokoušejí utratit více finančních prostředků, než kolika disponují, a proto se pokoušejí na finančním trhu dané prostředky získat.(12)

Proces financování na finančních trzích probíhá neustále a to buď přímo, polopřímo anebo nepřímě. Při přímé alokaci finančních zdrojů dochází k přímému pohybu volných finančních prostředků od přebytkového subjektu k deficitnímu tím, že deficitní subjekt emituje například investiční instrumenty, především cenné papíry. Jedná se o tzv. primární cenné papíry, za které obdrží od věřitele hotovost při přímém prodeji. Jedná se o nejjednodušší způsob přesunu volných finančních prostředků od investora k emitentovi. Polopřímé financování představuje mezistupeň mezi přímým a nepřímým

financováním. Přebytkové a deficitní jednotky se mezi sebou přímo nesetkávají, ale využívají služeb prostředníka, kterým je zpravidla specializovaná společnost, jenž poskytuje zdroj informací jak pro přebytkové tak pro deficitní subjekty.

Nepřímá alokace finančních zdrojů je daná rozhodující rolí zprostředkovatelů ve formě komerčních bank, pojišťoven, spořitelen, penzijních fondů nebo investičních společností. Tito zprostředkovatelé sami emitují své vlastní investiční instrumenty ve formě sekundárních investičních instrumentů, které nabízejí přebytkovým subjektům. Za získané finanční zdroje nakupují od deficitních subjektů jejich emitované investiční instrumenty. Nepřímá alokace finančních zdrojů umožňuje vysoce efektivní a likvidnější přesun volných finančních zdrojů od přebytkových subjektů k deficitním s minimálními informačními a transakčními náklady.(14)

1.2 Rozdělení finančních trhů

Dělení finančních trhů lze specifikovat dle různých hledisek. Používaná členění se mohou lišit podle času, četnosti, věcného předmětu obchodování nebo místa obchodu. Finanční trh lze však také rozdělit na privátní (tj. veškeré akce uskutečněné na finančním trhu jsou neveřejné a probíhají mezi oběma subjekty přímo) a zprostředkovatelský (v rámci transakce existuje třetí subjekt, který zajišťuje samotný přesun kapitálu). Nejčastějším a hlavně nejdůležitějším dělením finančního trhu je však hledisko délky splatnosti finančních instrumentů. Díky tomuto typu rozdělení lze odlišit trh peněžní a kapitálový.(12)

1.2.1 Peněžní trh

Základní vlastností peněžního trhu je krátkodobost prováděných finančních operací. Je to trh, kde se subjekty s přechodným přebytkem peněžních prostředků setkávají se subjekty, které vzhledem k jejich přechodnému nedostatku tyto prostředky poptávají. Peněžní trh je tvořen trhem krátkodobých úvěrů a trhem krátkodobých cenných papírů, splatných zpravidla do jednoho roku. Tyto nástroje se obecně vyznačují především poměrně malým rizikem, nižším výnosem, a co se týče cenných papírů, většinou i jejich poměrně vysokou likviditou. Jednou z hlavních funkcí peněžních trhů je financování

provozního kapitálu podniků a rovněž poskytování krátkodobých peněžních půjček, jak domácnostem, tak zejména vládám.(12)

1.2.2 Kapitálový trh

Segment kapitálových trhů zaznamenal masivní rozmach počátkem devadesátých let. Za tímto rozvojem stály především procesy internacionalizace, globalizace a s tím související deregulace, standardizace či sekuritizace. Je nutné připomenout, že za rozvojem kapitálových trhů nejen u nás, ale i ve světě, stojí konec studené války. V této době se do popředí dostává politika a kapitálové trhy v USA, popř. v Evropské unii. Pokračující rozvoj kapitálových trhů zapříčinil jak rozmach společného světového trhu, tak vznik trhů národních. Růst významu kapitálových trhů dokreslují různé studie kapitalizace cenných papírů. (4)

Pod kapitálovým trhem se skrývá trh dlouhodobých finančních instrumentů, u kterých je doba splatnosti delší než jeden rok. Díky délce splatnosti se za instrumenty kapitálového trhu považují akcie, dlouhodobé dluhopisy emitované státem, obchodními společnostmi nebo bankami. Díky této skutečnosti se kapitálový trh také nazývá trhem cenných papírů. Kromě akcií a dluhopisů jsou na kapitálových trzích obchodovány podílové listy, investiční certifikáty, hypoteční zástavní listy nebo finanční deriváty v podobě opcí, financial futures či swapů. Výše jejich výnosů je rozdílná, ale z obecného hlediska je s těmito investičními instrumenty spojeno vyšší riziko, a tím i vyšší výnos než u instrumentů peněžního trhu.(14)

Subjekty kapitálového trhu jsou jednotlivci, firmy, finanční instituce, státy, mezinárodní a nadnárodní instituce, které na trhu působí jako investoři, emitenti a finanční zprostředkovatelé.

Kapitálový trh dále můžeme rozdělit podle následujících hledisek:

1. Podle emise cenných papírů:

- Primární trh.
- Sekundární trh.

Funkce **primárního trhu** spočívá v získávání nových peněžních zdrojů z nově emitovaných cenných papírů. Emitenti získávají prodejem cenných papírů na primárním trhu dodatečný kapitál, přičemž se jedná o prvotní prodej investorům. Primární trh se také nazývá trhem nových cenných papírů. Na primárním trhu mohou být cenné papíry prodány pouze jednou.(12)

Na **sekundárním trhu** dochází k opětovnému prodeji cenných papírů. Samotné obchodování je zde rozsáhlejší, než na trhu primárním z prostého důvodu. Hlavní funkcí sekundárního trhu je stanovování tržní ceny cenných papírů a zajišťování jejich likvidity.

2. Podle způsobu obchodování:

- Veřejný trh.
- Neveřejný trh.

Na **veřejných trzích** se mohou zúčastnit všichni potenciální zájemci. Obchody jsou zprostředkovány organizátory trhu s cennými papíry, kteří zajišťují vypořádání všech transakcí za nejvyšší tržní cenu jak na trzích primárních tak i sekundárních.

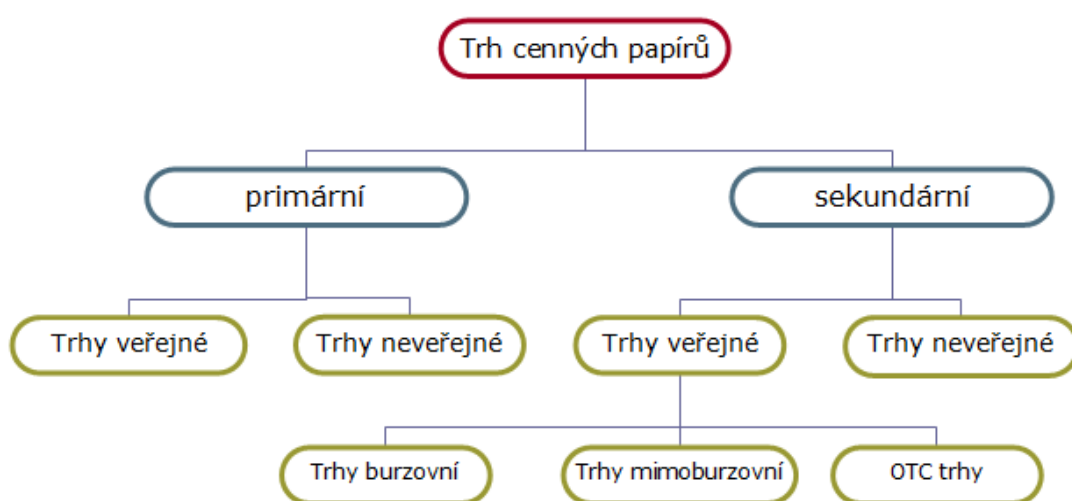
Veřejné trhy sekundárního trhu cenných papírů můžeme dále rozdělit na:

- **Trhy burzovní**, které reprezentují burzy cenných papírů. Burza cenných papírů je organizovaný trh, jehož činnost je upravena zákonem. V České republice existoval do konce roku 2008 jediný organizovaný subjekt – Burza cenných papírů Praha, a.s.
- **Organizované trhy mimoburzovní** jsou Českou národní bankou licencované a zákonem upravené organizace, které vykazují obdobnou činnost jako burzy cenných papírů, mnohdy ve významně větších objemech. Jediný mimoburzovní trh v České republice, RM-Systém, se 1. 12. 2008 přetransformoval na standardní burzu.
- **Neorganizované trhy** – tzv. „obchody přes přepážku“, neboli OTC trhy¹ jsou zpravidla ne tak silně regulované trhy, které jsou zprostředkovávány zejména bankami, nebo jinými institucionálními či soukromými obchodníky s cennými papíry.

¹ OTC – Over the counter markets

Na **neveřejných trzích** probíhají pouze tzv. smluvní obchody, při nichž bývají předmětné cenné papíry prodány, buď pouze jednomu, nebo několika investorům zároveň, a to na základě individuálních podmínek. Tyto obchody se uskutečňují, jak na trzích primárních, tak i sekundárních.(12)

Schéma kapitálových trhů:



Obrázek 1: Rozdělení finančního trhu

1.3 Predikce a akciové trhy

Je logické, že se samotným vznikem trhů s cennými papíry vznikla i snaha o vývoj jejich budoucích hodnot. K největšímu rozmachu metod akciové analýzy došlo zejména po druhé světové válce. Vedlo to ke vzniku specializovaných profesí – akciový analytik, bankovních oddělení pro analýzu cenných papírů – Research Departments a vznikaly dokonce specializované poradenské společnosti pro oblast investování, které se zabývaly analýzou akciových trhů – Moody's Investor Service, Standart & Poor.

Alfou a omegou obchodování na akciovém trhu je ve správný okamžik akcie levně koupit a ve správný okamžik je draze se ziskem prodat. Problémem je, že kdy nastane

ten správný okamžik nikdo jistě neví. Proto vznikly akciové analýzy, které se na základě různých faktorů snaží nalézt ty, které tržní cenu akcií nějakým způsobem ovlivňují. Cenový vývoj na akciovém trhu, jež je obecně vyjadřován indexy, je jedním z nejdůležitějších předstihových ukazatelů v konjunkturálním výzkumu pro indikaci hospodářských cyklů. Vývoj na akciovém trhu předbíhá skutečný vývoj HDP přibližně o 6 měsíců. Proto analýza a predikce vývoje akcií je důležitá pro pochopení a prognózy vývoje hospodářského produktu.(10)

1.4 Analýza akcie

Při rozhodování, kdy a do čeho vložit své volné finanční prostředky, si můžeme pomoci třemi typy analýz – technická, fundamentální a psychologická. Tyto analýzy se liší ve vnímání důležitosti údajů o investičních instrumentech. Každá z uvedených analýz má své odpůrce i zastánce. V této práci budou jako vstupní parametry prediktivního algoritmu použity s největší pravděpodobností indikátory technické analýzy (EMA 20, EMA50, EMA 200), nicméně stručně popíší i zbývající dvě.

1.4.1 Fundamentální analýza

Fundamentální analýza pracuje s fundamentálními daty, jako jsou zisky, tržby, očekávaný růst zisku a dividend, atd. Fundamentální analytici věří, že existují cenotvorné faktory, které determinují vnitřní cenu. Psychologická analýza se zajímá o účastníky finančního trhu a snaží se předpovědět jejich chování. Věřící, že lidé své chování opakují, a tak se snaží odpovédět na otázku, co budou dělat ostatní.(4)

Fundamentální analýza má četné příznivce i stoupence. Mezi obhájce této metody patří například nositel Nobelovy ceny profesor Loist, který uvádí: „*Fundamentální analýzy se snaží zjistit vliv podnikových a globálních faktorů na vývoj akciových kurzů a využít je pro ziskovou investorskou strategii. Vzdor problémům s oceňováním poskytuje fundamentální analýza hodnotné údaje o budoucím vývoji podniku. Fundamentální ukazatele patří k samozřejmému instrumentariu analytiků cenných papírů v praxi. Nové výzkumy ukázaly, že vyhodnocování ročních uzávěrek nelze ničím nahradit*“².

² Loistl, O.: Computergestütztes Wertpapiermanagement, str. 246.

Opačný názor na fundamentální analýzu má např. R.D. Edwards a J. Magge: „*Je nesmyslné chtít určovat vnitřní hodnotu akcií. Statistická data, která fundamentální analytici studují, jsou minulost. Burzovní trh se nezajímá o minulost, ale o přítomnost. Burzovní kurz je určován pouze nabídkou a poptávkou. Burzovní kurz nereflektuje rozdílné hodnocení vnitřních hodnot substančních analytiků. Kurz reflektuje naděje, obavy, domněnky a nálady - racionální a neracionální - tisíců potenciálních kupujících a prodávajících*“³.

Postup prognózy se děje ve třech krocích:

- 1) **Globální analýza** - zkoumá celkovou hospodářskou situaci odvětvové tendence.
- 2) **Odvětvová analýza** - zkoumá odvětvové tendence.
- 3) **Analýza konkrétní akcie** - zkoumá vnitřní hodnotu akcie.

Ad 1) Globální analýza

Prognózuje vývoj akciového trhu jako celku. Jejím cílem je odhadnout budoucí makroekonomické ukazatele a jejich vliv na akciové kurzy.

Nejdůležitější ukazatele, jimž se globální analýza věnuje, jsou:

1. **Hrubý domácí produkt** a míra jeho růstu - teorie říká, že vývoj na akciovém trhu předbíhá vývoj ekonomiky asi o šest měsíců. Důvodem je prozíravost investorů, kteří dokáží předvídat růst HDP. Tím se zpětně ovlivňuje i růst produktu, poněvadž díky růstu akciového trhu stoupá i reálný majetek akcionářů, kteří se stávají bohatšími a začínají více utrácet, čímž roste agregátní poptávka a zároveň roste i produkt. (1)
2. **Fiskální politika** - existence daní a byrokracie ovlivňuje akciové trhy velmi negativně. Daně firmám snižují zisky, a tím i schopnost vyplácet dividendy a podvazují růstové možnosti firmy a tudíž se pro potencionální investory stávají méně atraktivní. Pokud se tedy zvýší daňové sazby, akciový trh na to reaguje poklesem a naopak. Výjimkou zde jsou firmy, které profitují ze státních zakázek. (1)

³ Edwards, R.D. - Magge, J.: Technische Analyse von Aktientrends, str. 7 - 8.

3. **Monetární politika** - změna nabídky peněz v ekonomice je v neinverzním vztahu s akciovým trhem. Růst nabídky peněz díky efektu vyšší likvidity má přímý vliv na růst akciových kurzů, tomu napomáhá většinou i to, že s růstem peněžní nabídky jsou úrokové sazby nižší než při restriktivní měnové politice, jakou například prožívala Česká republika v letech 1997-1998. (1)
4. **Vliv úrokových sazeb** - změny úrokových sazeb jsou v inverzním vztahu s akciovými kurzy, protože růstem úrokových sazeb roste i nominální úrokové zhodnocení alternativních a relativně bezpečnějších investic např. obligací, či termínovaných vkladů a tím se snižuje poptávka na akciových trzích. (1)
5. **Inflace** - při stabilní ekonomické situaci má inflace na akciové kurzy neutrální vliv, protože akcie jsou podloženy reálným majetkem, který inflací neztrácí svoji hodnotu. Pokud však inflace neočekávaně vzroste, ekonomická nejistota se přelije i na akciový trh, což sebou přinese i relativní znehodnocení akciových kurzů. (1)
6. **Mezinárodní pohyb kapitálu** - zahraniční kapitál má na akciový trh velký vliv, který je tím větší, čím je domácí trh méně likvidní. Příliv dlouhodobého zahraničního kapitálu má na akciové kurzy jednoznačně pozitivní efekt, který je však v poslední době znehodnocován spekulativním kapitálem, který může naráz ihned odplynout pryč a přivést akciový trh a celou zemi do krátkodobé krize (Mexiko 1995, jihovýchodní Asie 1997-1998, Brazílie 1999). (1)
7. **Ekonomické a politické šoky** mají na ekonomiku včetně akciového trhu silně negativní vliv z důvodu destabilizace ekonomické, potažmo i společenské situace. Tyto šoky mohou ekonomiku uvrhnout do dlouhodobé recese. Mezi nejvýznamnější šoky posledních desetiletí patří: válka v Perském zálivu 1991-1992, válka na Balkáně, 1992-2000 - politické; Ropná krize 1973 a 1979, Světová Finanční krize 2008 – ekonomické. (1)

Ad 2) **Odvětvová analýza**

Prognózuje vývoj v jednotlivých odvětvích a analyzuje jejich rozdílné charakteristiky. U každého odvětví se zjišťuje zejména nákladová stránka výroby, míra zisku, exportní schopnosti, inovační a technologická očekávání a jejich vliv na akciové kurzy.

Jednotlivá odvětví rozlišujeme podle toho, jak reagují na změnu produktu a agregátní poptávky:

- **Cyklická odvětví** - kurzy akcií se pohybují s hospodářským cyklem. Při expanzi akciové kurzy rostou a naopak při recesi klesají (např. automobilový průmysl). (1)
- **Neutrální odvětví** - kurzy akcií nereagují na hospodářský cyklus, protože poptávka po produktech odvětví je relativně stabilní (např. potravinářský průmysl). (1)
- **Anticyklická odvětví** - kurzy akcií se pohybují inverzně s hospodářským cyklem. Zhoršení ekonomické situace vede k přesunu poptávky od nákladnějších a kvalitnějších produktů k produktům jednodušším a podřadnějším (např. televizní společnosti). (1)

Ad 3) **Analýza konkrétních společností**

Fundamentální analýza předpokládá, že každá akcie má svou vnitřní hodnotu, která vychází z historických dat z hospodaření společnosti a aktuální kurz akcie se pohybuje kolem této vnitřní hodnoty. Tato analýza se skládá ze dvou postupných kroků, nejprve za použití finančních analýz se snaží zjistit vnitřní hodnotu a prognózovat budoucí vývoj společnosti, a pak tyto údaje srovnává s aktuálním kurzem a snaží se zjišťovat, zda je akcie podhodnocena (impuls k nákupu) nebo nadhodnocena (impuls k prodeji), což udržuje kurz kolem vnitřní hodnoty. (1)

Na zjišťování vnitřní hodnoty akcie existuje množství různých postupů, modelů a výpočtů různých hodnot (dividendový diskontní model, ziskové modely - ratio P/E, cash-flow model, bilanční model), které jsou však většinou složité, časově náročné a jejich vypovídací schopnost je dosti pochybná. (1)

Jediné s čím lze plně souhlasit je, dle mého názoru, užitečnost likvidační hodnoty. Likvidační hodnota je množství peněz, které zůstane firmě i po ukončení podnikání po prodeji všech svých aktiv a vypořádání všech svých závazků a zbytek peněz se rozdělí mezi akcionáře. Likvidační hodnotu na akcii můžeme charakterizovat jako spodní hranici tržní ceny akcie a v případě poklesu pod tuto hodnotu dochází k pokusům o převzetí firmy za účelem její likvidace, což přinese spekulantovi zisk. (1)

1.4.2 Psychologická analýza

Oproti fundamentálnímu nebo technickému přístupu je velmi odlišná, jelikož nezkoumá daný cenný papír (a co je za ním), ostatní fundamenty, ani grafy, ale lidské chování při investování. Jelikož je investování kolektivní záležitost, stává se stěžejním bodem analýzy psychologie davu a impulsy k masovému nakupování/prodávání. (11)

Dobré základy k poznání psychologie davu položil francouzský sociolog, psycholog, matematik a lékař Gustave Le Bon. Základem jeho vysvětlení je fakt, že vlastnosti davu nejsou dány charakterovými vlastnostmi jednotlivých členů davu, ale vytváří se nové. Jedinec, který vstoupí do davu, ztrácí veškeré racionální myšlení a funguje především na bázi svých pudů a citů. Dav proto většinou postrádá veškeré racionální myšlení. Podle Le Bona ovšem existují určité prvky, které chování davu určitým způsobem ovlivňují. Těmito prvky jsou například vzdělání, výchova, rasa apod. Stavebním kamenem každého psychologického analytika je poznání chování davu. Dav se většinou neřídí zkušenostmi z minula, ale často dává na „horké tipy“ známých analytiků a investorů. Stává se tedy velmi snadno zmanipulovatelným a těžko se jeho chování odhaduje. Na koncepci Le Bonova davu postavilo své investiční strategie mnoho investorů. Ti nejlepší své koncepce obchodování podle psychologické analýzy zformulovali a předali dále. Patří sem zejména koncepce známého André Kostolanyho, Johna Maynarda Keynesa, George Drasnara, Iry Epsteinové a Davida Garfielda a řadíme sem také teorii spekulativních bublin. (11)

1.4.3 Teorie spekulativních bublin

Tato teorie vysvětluje odchylky od správného kurzu opět podobným způsobem jako předchozí dvě koncepce. U spekulativních bublin je ovšem rozdíl v tom, že jsou vždy způsobeny nadměrnou reakcí investorů na určitou událost. V davu totiž jednoduše dochází k davové psychóze a investoři pak i sebemenší podnět vnímají daleko intenzivněji (jednoduše, i prakticky nevýznamný podnět může být vnímán přehnaně). Prakticky dochází k růstu bubliny díky vlně optimismu v davu. Bublina se neustále zvětšuje do doby, dokud dav nezlomí určitý podnět, který změní jejich očekávání v pád. Bublina potom velmi rychle praskne a dochází k velmi rychlému pádu kurzu, který může být velmi nebezpečný. Tento jev již zlomil vaz zástupu velmi zkušených investorů. Jeho nebezpečí spočívá v tom, že se bublina velmi obtížně pozná, nástup do ní je velmi lákavý a pozdější pád velmi rychlý. (11)

1.4.4 Technická analýza

Technická analýza je postavena na analýze publikovaných tržních dat, kterými jsou kurzy nebo objemy obchodů. Dá se využít jak na trzích s cennými papíry, tak třeba i na komoditních trzích. Hlavním cílem technických analytiků je prognózování krátkodobých pohybů akcií nebo akciových indexů. Technická analýza se snaží předpovědět pohyby kurzu jednotlivých akcií pomocí analýzy minulého vývoje kurzů, objemu obchodů a minulých trendů. Je reakcí na kritiku fundamentální analýzy, její neschopnosti určit vnitřní hodnotu a zachytit kurzotvorné změny. Nedůvěřuje pramenům, podle kterých tvoří fundamentalisté své předpovědi, a vytýká jim časovou náročnost. Proto se místo toho zaměřují na změny tržních cen jako indikátory nabídky a poptávky. Technická analýza je proto založena na předpokladu, že všechno, co potřebuje investor vědět, je obsaženo v tržních cenách. (1) (11)

Metod technické analýzy je mnoho, ale všechny se odvíjejí od pěti základních principů:

- 1) Kurzy akcií jsou určovány pouze nabídkou a poptávkou.
- 2) Nabídka a poptávka je ovlivňována racionálními i iracionálními činiteli.

- 3) Kurzy akcií se mají tendenci pohybovat se v trendech po významné časové úseky.
- 4) Změny v trendech jsou způsobovány posuny poptávky a nabídky a lze je včas identifikovat studiem historických cen a objemu obchodů.
- 5) Mnoho grafických formací v grafech a diagramech má tendenci se opakovat, což umožňuje prognózovat budoucí vývoj kurzů akcií.

Dnešní nástroje technické analýzy obsahují nepřehledné množství metod a postupů, které se však dají rozdělit do dvou zcela základních skupin. Jsou to metody

- Grafické – hledají snadno rozpoznatelné vzory – opakující se jevy v grafu ceny akcií nebo objemu prodeje.
- Technické indikátory – jsou to matematické funkce využívající časový průběh změn, tzn. časové řady.

1.4.5 Technické indikátory

Technická analýza disponuje nepřehledným množstvím indikátorů. Nejčastější metodou technické analýzy je studie grafů. Zkoumají se aktuální i minulé ceny a hodnoty od nich odvozené pomocí různých vzorců a výpočtů. Tyto ukazatele (indikátory) nám mohou poskytovat informace o objemu obchodů (volume), dočasné tzv. přeprodanosti resp. překoupenosti trhu a hybnosti trhu (cenové oscilátory, momentum). Nejjednoduššími technickými ukazateli jsou však trendové přímky, hladina podpory, hladina rezistence a klouzavé průměry. (9)

Pro řešení zadání diplomové práce budu pracovat s tzv. klouzavými průměry. Tyto indikátory patří do skupiny indikátorů opožděných. Druhou skupinou jsou pak indikátory vedoucí.

Klouzavý průměr (MA - Moving Average)

Mezi nejoblíbenější indikátory patří klouzavé průměry a je na nich založeno mnoho metod. Zobrazují průměrnou hodnotu ve zvolené časové periodě. Zavěsí se na nově nastoupený trend, následují ho, a pokud se změní, změní se i směr klouzavého průměru. Nejčastějším vyhodnocováním signálů je sledování situace, kdy se indikátor protne s cenou.

Pokud indikátor protne cenu zezdola nahoru, lze to chápat jako signál růstu. Analogicky pak signálem sestupu je protnutí ceny indikátorem shora dolů.

V této práci bude jako vstupní parametr sloužit mimo jiné exponenciální klouzavý průměr (*EMA – Exponential Moving Average*).

SMA – Simple Moving Average

Jednoduchý klouzavý průměr je nazýván jednoduchým z prostého důvodu. Nejedná se totiž o nic jiného, než o průměr hodnoty (open, high, low nebo close) za dané časové období. Neobsahuje žádnou vyhlazovací složku a je tedy velmi náchylný ke změně, aniž by došlo ke změně na trhu. Jinými slovy, změna klouzavého průměru nemusí reflektovat změnu trendu nebo změnu na trhu. Je to způsobeno tím, že pokud je nejstarší hodnota, ze které se průměr počítá, příliš vysoká nebo nízká, je v dalším dni z výpočtu vypuštěna a nahrazena cenou nejnovější. Vypuštění například vysoké nejstarší ceny má za následek snížení průměru za období, přestože žádná aktuální změna se neodehrála. Opačně to samozřejmě platí pro vypuštění příliš nízké ceny. To má za následek zvýšení klouzavého průměru, který opět nereflektuje aktuální dění s akcií.

$$SMA_t = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i}$$

EMA – Exponential Moving Average

Exponenciální klouzavý průměr dává největší váhu nejnovějším hodnotám, ale zároveň bere v úvahu všechny minulé data a to i ty, které už nespádají do uvažované časové periody. Váhy rostou exponenciálně, přičemž váha přiřazovaná nejnovější hodnotě je označovaná jako vyhlazovací faktor (smoothing factor) nebo též exponenciální procento. Tento vyhlazovací faktor je určen na základě délky periody MA, nebo může být stanoven primárně, z čehož následně vyplýne vhodná délka periody.(10)

EMA je tedy daleko vhodnější k použití pro predikci, jelikož není zatížen stejným neduhem jako SMA. Nicméně vzhledem k rekurzivnímu způsobu výpočtu je nutné získat počáteční hodnotu indikátoru. A právě zde nalezne využití SMA, který je pro započetí výpočtu EMA dostačující. Vzorec pro výpočet je následující:

$$EMA_t = EMA_{t-1} + \alpha(P_t - EMA_{t-1})$$

Vytknutím lze vzorec přepsat takto:

$$EMA_t = EMA_{t-1} + \alpha P_t (1 - \alpha) - EMA_{t-1}$$

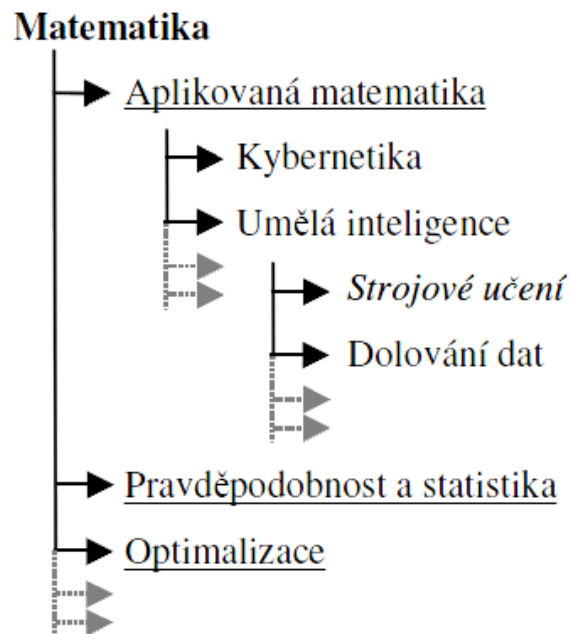
Kde $\alpha = \frac{2}{n+1}$ je vyrovnávací konstanta (vyhlazovací faktor).

1.5 Umělá inteligence a strojové učení

Přestože je nosná část této práce, Bayesovské klasifikace, metodou pravděpodobnostní, je spojována s umělou inteligencí. Bayesovské klasifikátory jsou intenzivně studovány v souvislosti se strojovým učením a velmi často jsou využívány v genetických algoritmech.

1.5.1 Zařazení strojového učení

Strojové učení (SU) je českým ekvivalentem anglického termínu *Machina Learning* (ML). Jedná se o vědní disciplínu zabývající se algoritmy a programy, které umožňují strojům proces *učení*. Samotné hierarchické zařazení SU jako kategorie mezi ostatními vědními disciplínami není jednoznačné. Důvodem je jejich vzájemné prolínání se a doplňování. Následující schéma proto není jediným možným způsobem, kam SU zařadit. (6)



Obrázek 2: Zařazení oboru Strojové učení dle ing. P. Honzíka, Ph.D.(6)

Mezi obory, které významným způsobem zasahují do SU, a přitom jsou hierarchicky v odlišné oblasti, náleží zejména *statistika* a *optimalizace*. Statistika nabízí silný matematický aparát využívaný zejména ve fázi předzpracování dat a při vyhodnocování modelu (průběžném i konečném)⁹. Termín model lze z pohledu statistiky nahradit termínem *hypotéza* (model je ve své podstatě hypotézou o vztazích a vazbách mezi vstupními a výstupními veličinami v modelovaném procesu či objektu). Optimalizace se zabývá algoritmy, které vedou k nalezení extrému funkce. Tento obor tedy na model pohlíží jako na *funkci*. Cílem optimalizace v SU je nastavení parametru modelu tak, aby při daných vstupních hodnotách byla minimalizována odchylka mezi výstupem požadovaným a získaným z modelu(6).

1.5.2 Umělá inteligence

Samotnou inteligenci stručně charakterizuje Vladimír Mařík:

„Intelligence je vlastností některých živých organismů, která jim dává v přírodě mimořádné postavení. Vznikla a vyvíjela se v průběhu dlouhého vývoje. Dnes umožňuje

některým živým organismům efektivně reagovat na složité projevy prostředí a aktivně je využívat ve svůj prospěch, k dosažení svých cílů.“⁴

Jak je ze schématu na obr. 2 patrné, umělá inteligence (UI), ang. Artificial Intelligence (AI), je oboru SU nadřazená. Marvin Minský definoval UI jako vědu, která se zabývá tím, jak přimět stroje projevovat se takovým chováním, které by v případě člověka vykazovalo potřebu inteligence. Existuje tzv. Turingův test, který zjednodušeně spočívá v tom, že člověk klade otázky (např. písemnou formou) a dostává odpovědi. Na základě těchto odpovědí se rozhoduje, jestli komunikuje s člověkem nebo strojem. Pokud stroj při tomto testu oklame člověka, projevuje se inteligentně. Je to však opravdu důkaz existence umělé inteligence? Stroj mohl nalézt odpovědi v nějaké databázi, přitom ve skutečnosti vůbec nechápe jejich smysl. Pak hovoříme o tzv. *slabé UI*. *Silná UI* naproti tomu chápe obsah podobně jako člověk. Nedostatečnost Turingova testu na dokázání silné UI byla provedena pomocí tzv. čínského pokoje. Osoba neznalá čínštiny, jen na základě určitých pravidel, korektně odpoví na otázky v tomto exotickém jazyce. Princip tohoto důkazu přirovnal Stanisław Lem ke skládání puzzle obrázku. Podstata pokusu spočívá v tom, že dílky jsou obráceny rubem nahoru, není tedy možné puzzle sestavit na základě obrázku, jehož vzhled je vodítkem při skládání. Přesto je možné v případě, že je každý dílek unikátní (vhodný formální popis jazyka), poskládat puzzle správně tak, aby obrazec (nebo např. text) na lícové straně dával smysl. Tím byla prokázána nedostatečnost Turingova testu na dokázání silné UI. Vývoj UI ovlivňuje nejen akademická obec, ale dokonce celá oblast literatury označovaná jako science fiction si pohrává s prognostickými myšlenkami o možném vývoji a vlivu UI. Prognózy některých spisovatelů jsou brány velice vážně (A.C.Clark, Stanisław Lem, ...). Smyslem těchto příkladů bylo vysvětlit, že UI se jako obor nezabývá pouze přímou aplikací matematických modelů, ale také filosofickými a etickými důsledky plynoucími z jejich použití. Skutečná realizace však nakonec končí u algoritmu a konkrétních matematických modelů (6).

⁴ MAŘÍK, V. *Umělá inteligence 1*. 1. vyd. Praha: Academia, 1993. 264 s. ISBN 80-200-0496-3. Kapitola 1, Úvod, s. 15–16.

1.5.3 Strojové učení

Strojové učení je tedy nauka o algoritmech, které umožňují učení umělých objektů. *Učením se* je rozuměno automatické zlepšování se na základě zkušeností. SU se tedy orientuje na matematický aparát, na jehož základě dochází k technické realizaci cílů definovaných v umělé inteligenci. Skutečné meze těchto algoritmů ve vztahu k člověku a jeho chápání včetně ostatních filosofických úvah do SU nespadají.(6)

1.5.4 Statistika

Statistika je s algoritmy strojového učení spjata velice úzce. Řada postupů používaných v SU vychází přímo ze statistických metod. Zcela zásadní rozdíl mezi SU a statistikou spočívá v tom, že podstatou SU je *vytvořit model (hypotézu)*, který bude predikovat a ve finále skutečně rozhodovat, přičemž proces učení nemusí vést při opakovaném výpočtu ke stejným výsledkům, tedy stejným modelům. Cílem je nastavit model co nejlépe. Statistika je matematický obor, který se zabývá popisem (charakteristiky datových souboru), indukci (předpoklady o základním souboru ze vzorku), porovnáváním a testováním (posuzuje hladinu významnosti hypotéz na základě předpokladu). Co do číselných výpočtů je na rozdíl od SU jednoznačná. Umožňuje tak absolutní hodnocení, na jehož základě lze provádět různá srovnání. Zjednodušeně řečeno, SU vytváří hypotézy na základě velice pestré škály názoru, nápadu a postupu. K posouzení míry zdařilosti této činnosti (ať už během procesu učení nebo po jeho ukončení) se pak obrací na statistiku jako na metodologii garantující korektnost při absolutním posuzování kvality vytvořených modelů(6).

1.5.5 Optimalizace

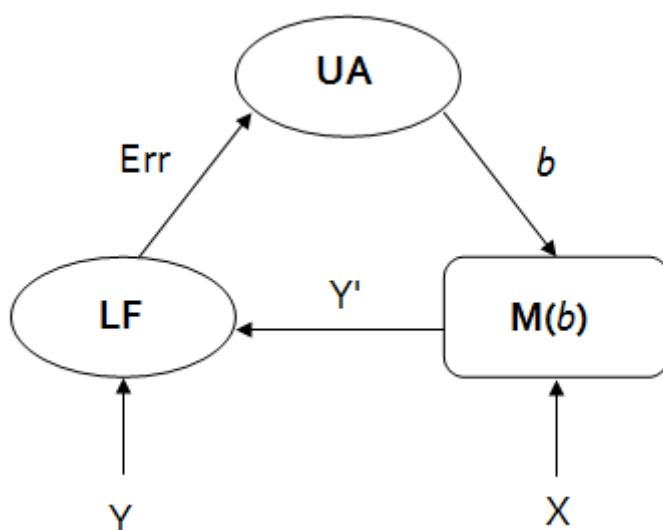
Optimalizace je poslední oblastí, která hraje v SU zásadní roli. SU zahrnuje mnoho různých modelů, ve kterých lze uložit znalost v podobě jejich struktury a parametru. Nastavení hodnot parametru závisí na vstupních informacích, tedy trénovacích datech. Postup, jak nastavit model na základě dat, se nazývá optimalizace (spočívá v

minimalizaci odchylek rozdílů výstupních hodnot modelu a požadovaných výstupních hodnot).

1.6 Komponenty procesu modelování

1.6.1 Základní schéma

Schéma na obrázku 3 znázorňuje objekty a funkce, které dohromady tvoří základní komponenty procesu modelování z pohledu strojového učení. Je tvořeno vstupní veličinou X , výstupní veličinou Y , predikovanou výstupní veličinou Y' , modelem s parametry $M(b)$, chybovou funkcí LF a optimalizačním (učícím) algoritmem UA .(6)

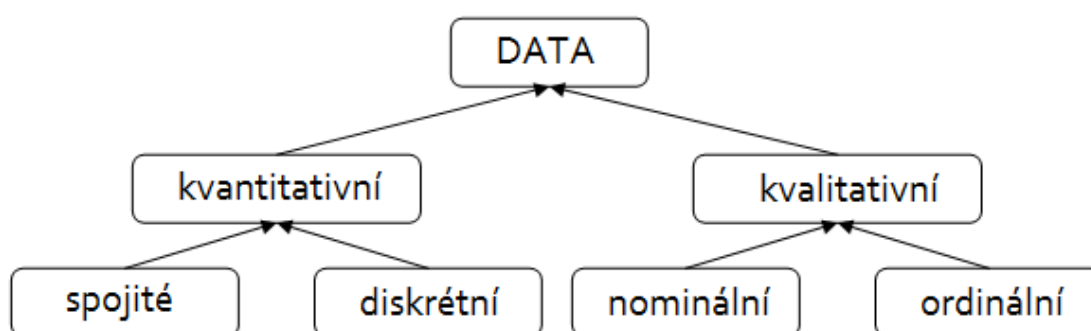


Obrázek 3: Model učení s učitelem (6)

Jeden cyklus procesu učení s učitelem probíhá následovně: postupně jsou předkládána vstupní data X modelu $M(b)$; výstup modelu Y je porovnáván s požadovaným výstupem Y ; míru vzniklé odlišnosti kvantifikuje chybová funkce LF (ang. loss function) na tzv. chybu Err ; optimalizační (učící) algoritmus UA upravuje parametry modelu b na základě informace o chybě Err tak, aby její hodnota byla v příštím cyklu co nejmenší.(6)

1.6.2 Veličiny

Vstupní, výstupní i predikované veličiny (X , Y , Y') mohou být různého typu. Jejich základní dělení znázorňuje obrázek 4. Charakteristické pro kvantitativní veličinu je skutečnost, že její číselná hodnota má nějakou jednotku (metry, procenta, počet lidí v tramvaji,...) a vyjadřuje množství (kvantum) v dané jednotce. Mezi kvantitativními daty pak existuje operace sčítání a odečítání. Kvantitativní veličiny se dále dělí na spojité a diskrétní (obdoba analogové a digitální).(6)



Obrázek 4: Typy veličin (6)

Kvalitativní veličiny nenesou číselnou informaci. Dělí se na ordinální a nominální. Prvky v ordinální veličině lze seřadit, ve smyslu orientace je mezi nimi možno definovat relace $<$, $>$, $=$. Často však relace závisejí na kontextu, ve kterém jsou data použita. Například vztaženo ke vzdálenosti od slunce lze množinu planet $\{Merkur, Venuše, Země, Mars, Jupiter, Saturn, Uran, Neptun\}$ považovat za ordinální. Neplatí však, že by vzdálenost mezi Zemí a Marsem byla stejná jako mezi Marsem a Jupiterem. Pokud bychom se zabývali hmotností planet, pořadí by bylo odlišné. Jestliže kontext neznáme nebo dokonce neexistuje, na pořadí prvku v množině nezáleží. Taková veličina je nominální.(6) (8)

1.6.3 Modely

Podle výše zmíněného dělení veličin na kvantitativní a kvalitativní se dělí i modely podle typu výstupní veličiny na *modely regresní* – výstup je kvantitativní, a *klasifikátory* – výstup je kvalitativní. Běžná je kombinace obou typu modelu dohromady, tzv. regresní klasifikátory. Jejich princip spočívá v tom, že kvalitativní výstupní veličina je nahrazena čísly, která jsou v regresním modelu použita jako kvantitativní veličina. Na spojitém výstupu regresního modelu je následně stanovena tzv. kritická hodnota (práh, mezní hodnota), na jejímž základě (překročení, nedosažení) je klasifikováno. Typická je tato substituce u binárních dat (muž/žena, ano/ne – nahrazeno čísly 0/1 nebo -1/1).(8)

1.7 Chybové funkce

Vstupy chybové funkce jsou požadovaná výstupní hodnota Y (G) a výstupní hodnota modelu \hat{Y} (\hat{G}). Cílem je kvantifikovat míru chyby modelu. Pokud platí rovnost obou výstupních veličin ($Y=\hat{Y}$, $G=\hat{G}$), je chyba nulová. V případě jejich odlišnosti chybu jednoznačně kvantifikuje chybová funkce LF . Pokud je výstupní veličina kvalitativní, používají se k určení chyby tzv. *nákladové matice*. Protože chybová funkce představuje zpětnou vazbu v optimalizačním procesu, je její korektní stanovení důležité. (8)

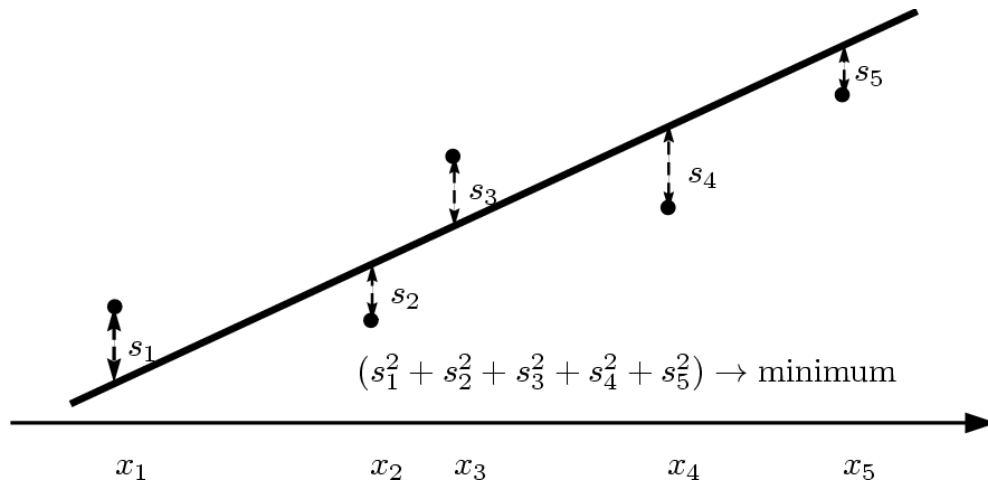
V případě Bayesovského modelu umělé inteligence, potažmo strojového učení založeného na bázi Bayesovské prediktivní (nebo lépe pravděpodobnostní) metodě je nejběžnější chybovou funkcí Metoda nejmenších čtverců, Minimum mean square error estimation – MSE. Přestože se v praktické části, potažmo ve zdrojovém kódu neobjeví, považují ji za část hodnou pozornosti. Pojednání o této metodě může být základem pro případnou budoucí implementaci do zjednodušeného prediktivního modelu.

1.7.1 Metoda nejmenších čtverců

Metoda nejmenších čtverců (MSE) je aproximační metoda, která spočívá v tom, že hledáme takové parametry zvolené funkce, pro které je součet čtverců odchylek vypočtených hodnot od hodnot naměřených minimální. Výpočet chyby Err podle uvedené definice vyjadřuje následující vztah:

$$Err = \sum_{i=1}^N [y_i - f(x_i, \mathbf{b})]^2$$

kde y je požadovaná výstupní hodnota, x je hodnota vstupní veličiny a \mathbf{b} je vektor parametru modelu $f(x, \mathbf{b})$. (7)



Obrázek 5: Metoda nejmenších čtverců (7).

1.7.2 Lineární MSE

Lineární MSE se týká regresních modelů, které jsou lineární ve svých parametrech \mathbf{b} , jak ukazuje následující rovnice

$$f(x; \mathbf{b}) = b_0 + b_1 f_1(x_1) + \dots + b_n f_n(x_n)$$

Hlavní předností modelu lineárních v parametrech je skutečnost, že pro nalezení optimálních hodnot jejich parametru zpravidla existují známá analytická řešení. Pokud je to možné, bývají z tohoto důvodu linearizovány i modely nelineární. To má většinou za následek, že rozložení chyby měření již není normální, což je implicitní předpoklad pro použití lineární MSE. Navzdory tomu však bývají výsledky takto upravených závislostí uspokojivě přesné. Nevýhodou modelu lineárních v parametrech je problematika aproximace složitějších závislostí a citlivost vůči prvkům, které jsou např. vinou chyby měření výrazně vzdáleny od prvku ostatních (ang. outliers). Pokud je

použita linearizace složitější závislosti, je třeba brát zřetel na interval, ve kterém toto zjednodušení platí; závislost nelze extrapolovat.(8)

1.7.3 Váhová MSE

V určitých případech může mít stejná odchylka při odlišné vstupní hodnotě různou váhu. Příkladem jest např. studie vztahu mezi předpokládanou a skutečnou cenou stavebního projektu. Stejná odchylka v malém projektu má odlišnou váhu než v projektu rozsáhlém. Typickou ukázkou je také rozdílný rozptyl chyby měření při jednotném rozsahu a různě velkými naměřenými hodnotami. Ukázkou upravené chybové funkce je následující vztah: (8)

$$Err = \sum_{i=1}^N [y_i - f(x_i, b)]^2 \cdot \frac{1}{x_i^2}$$

Váhovou MSE lze použít v lineárních i nelineárních modelech, jde pouze o modifikaci chybové funkce.

1.7.4 Přeučení modelu

Jedním z problémů při učení modelu na dostupných datech je tzv. *přeučení modelu*. Tento stav spočívá v tom, že se model naučí i zdánlivé, ve skutečnosti neexistující souvislosti. Důvodem může být příliš mnoho parametru modelu (Occamovo ostří) nebo zašuměnost a chyby v trénovacích datech. Vzniklý efekt je patrný z obr. 1.4. V uvedeném příkladu je vytvářen rozhodovací strom. Postupným přidáváním dalších uzlů je možno zvyšovat přesnost klasifikace na trénovacích datech. Z grafu je patrné, že od určitého počtu uzlů se chyba na testovacích datech začala zvětšovat. Příčinou je skutečnost, že byl model na

Trénovacích datech přeučen, tzn. naučil se závislosti, které ve skutečnosti neexistují a které ve svém důsledku vedou ke zhoršení přesnosti predikce. Vyhnout se přeučení je možno například Occamovým ostřím, omezeným prostorem hypotéz nebo preferencí jednoduchých hypotéz. (8)

1.8 Thomas Bayes (1702-1761)

Anglický duchovní a matematik, Thomas Bayes se narodil roku 1702 v Londýně. Bayes byl roku 1742 zvolen členem Royal Society, navzdory tomu, že v oblasti matematiky publikoval za svého života pouze jednu práci, a to navíc anonymně. Šlo o *Introduction to the Doctrine of Fluxions* (1736 – Úvod do nauky o infinitesimálním počtu; na vysvětlenou uveďme, že termínem „fluxional calculus“ Isaac Newton označoval infinitesimální počet). Bayes touto svou prací ostře nesouhlasně reagoval na spis *The Analyst* (1736), ve kterém irský teolog a filozof, biskup George Berkeley (1685 – 1753), napadl logické základy Newtonova infinitesimálního počtu. Druhým Bayesovým spisem, jenž vyšel za jeho života, byl *Divine Benevolence* (1731 – Božská dobrotivost).

Bayesova teorie pravděpodobnosti byla uvedena v jeho práci *Essay towards solving a problem in the doctrine of chances* (Esej směřovaná k řešení problému ve vědě o náhodě), publikované ve *Philosophical Transactions of the Royal Society of London* (Filozofická pojednání londýnské Královské společnosti) v roce 1764 – tedy již po Bayesově smrti. Bayesovu práci zaslal do Royal Society jeho přítel Richard Price, který ji našel v písemnostech pozůstalosti. Bayes se touto svou prací stal jedním ze zakladatelů induktivní statistiky (statistická analýza opírající se o vzorky tvořící malé podíly základního statistického souboru).(9)

1.8.1 Bayesův vzorec

V době, kdy počet pravděpodobnosti byl víceméně ve svých počátcích, Bayes se již zabýval problémem, jak by bylo možné uvádět pozdější zkušenosti do souladu s apriorními předpoklady, resp. jak takové zkušenosti změni následné hodnocení situace proti apriorním předpokladům. Šlo tedy do jisté míry o dynamické ověřování hypotéz, včetně korektury těchto hypotéz. (9)

V teorii pravděpodobnosti a v jejích aplikacích, ukazuje Bayesův teorém jak určit inverzní pravděpodobnosti: známe-li pravděpodobnost B podmíněnou událostí A, jaká je pravděpodobnost A podmíněná událostí B? Odpovědi na tuto otázku lze nalézt, je ale potřeba zahrnout takzvanou “apriorní“ neboli nepodmíněnou pravděpodobnost A a B. (9)

Tento teorém je také často nazýván Bayesův zákon či Bayesovo pravidlo. Bayesův teorém vyjadřuje podmíněnou pravděpodobnost neboli aposteriorní pravděpodobnost hypotézy H (jde o pravděpodobnost, u které víme, že E jistě platí) a apriorní pravděpodobnost hypotézy H , apriorní pravděpodobnost E a aposteriorní pravděpodobnost E je-li dáno H . Toto naznačuje, že jistota má potvrzující efekt, pokud je pravděpodobnější H než $\neg H$. (9)

Bayesův teorém je platný v běžných výkladech pravděpodobnosti a je obecně používán ve vědě a inženýrství a je intenzivně studován v souvislosti se strojovým učením a systémy pro dobývání znalostí. Nicméně, existuje spor mezi statistiky ohledně otázky, zda je možné Bayesův teorém použít k řešení všech otázek okolo problematiky inverzní pravděpodobnosti.(3)

Hlavní myšlenkou je, že pravděpodobnost události A podmíněné událostí B (například pravděpodobnost, že jedinec má rakovinu prsu podmíněná událostí, že byl pozitivně testován na mamografu) je závislá nejenom na vztahu mezi A a B , ale také na marginální pravděpodobnosti výskytu jevu A a jevu B .(2)

Například, je známo, že mamograf je spolehlivý na 95 procent. Tento výsledek může být dán 5 % chybných pozitivních testů nebo 5 % chybných negativních testů, a nebo může jít o mix pozitivních a negativních testů. Bayesův teorém umožňuje spočítat pravděpodobnost rakoviny, pokud výsledek mamografu byl pozitivní, pro všechny tři zmíněné situace. Pravděpodobnost pozitivního výsledku mamografu bude odlišná pro všechny tři případy. Ve zmíněném příkladě je znázorněn velký praktický dopad, kterého si stojí za to povšimnout: jestliže šance na pozitivní výsledek mamografu na rakovinu je řekněme 5 %, poté podmíněná pravděpodobnost, že osoba s pozitivním testem, skutečně má rakovinu je docela malá, poněvadž marginální pravděpodobnost tohoto typu rakoviny se blíží 1 %. Pravděpodobnost pozitivního testu je tedy 5 krát větší než pravděpodobnost skutečného onemocnění rakovinou. Rovněž můžeme odvodit, že podmíněná pravděpodobnost, kdy pozitivní výsledek mamografu naznačuje rakovinu prsu, je nanejvýše 20 %. Mohlo by to být i méně, jestliže je pravděpodobnost onemocnění rakovinou podmíněná tím, že výsledek mamografu je pozitivní, není 100%. Tento příklad ukazuje hodnotu správně pochopeného a aplikovaného Bayesova matematického teorému. (2)

1.8.2 Základní formulace Bayesova vzorce

Využívá se **apriorní pravděpodobnosti** ($p(H_k)$), která odpovídá počáteční pravděpodobnosti, a **aposteriorní pravděpodobnosti** ($p(H_k|D)$), která je vypočtená až při vlastní analýze a udává pravděpodobnost hypotézy. Principiálně je tento model blízký lidskému uvažování. (11)

$P(D)$ pravděpodobnost jevu D

$P(H_k|D)$ pravděpodobnost jevu H_k pokud víme, že nastal jev D

$$P(H_k | D) = \frac{P(D | H_k) P(H_k)}{\sum_{i=1}^k P(D | H_i) P(H_i)}$$

Po zjednodušení dostaneme

$$P(H_k | D) = \frac{P(D | H_k) P(H_k)}{P(D)}$$

H_k je **hypotéza** – například předpoklad určitého jevu (nemoc A, B nebo C)

D je realita, data, konkrétní měření

$P(H_k)$ je **apriorní pravděpodobnost**

$P(D|H_k)$ je **věrohodnost**, tedy pravděpodobnost daných výsledků za předpokladu například konkrétní nemoci – podmíněná hustota pravděpodobnosti udávající rozložení dat D uvnitř jednotlivých tříd.

$P(D)$ pravděpodobnost nastolení měření (dat, údajů) D , **evidence**

$P(H_k | D)$ je **aposteriorní pravděpodobnost** platnosti konkrétní hypotézy

Nechť je dán jistý vzorek dat $D = (d_1, \dots, d_n)$, který má být zařazen do jedné z tříd C_1, \dots, C_m . Zařadíme jej do třídy C_i , pro kterou platí:

$$P(C_i | D) \text{ je maximální}$$

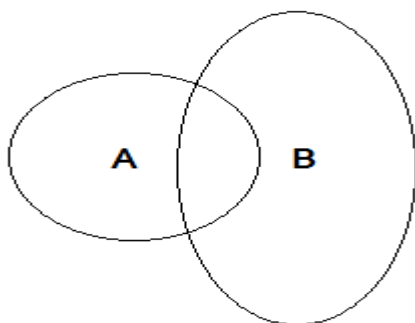
$$\text{Protože } P(H_k | D) = \frac{P(D | H_k) P(H_k)}{P(D)}$$

kde $P(D)$ je konstantní, hledáme maximální $P(C_i | D) P(C_i)$

Nejpravděpodobnější hypotéza je ta s největší aposteriorní pravděpodobností. Z ní se provádí další odhad. Výhodou je, že tato metoda využívá intuitivní řešení, které se podobá lidskému rozhodování a je možné kombinovat předchozí pravděpodobnost s pozorovanými daty.(3)

1.9 Bayesův vzorec – základní vztahy

Vztah pro slučitelné pravděpodobnosti, tedy pro jevy A a B, které se mohou vyskytovat současně.

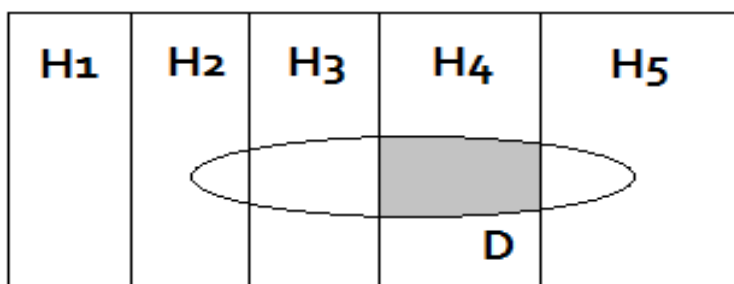


$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Pokud pro všechna $i, j; i \neq j$ platí $H_i \cap H_j = \{ \emptyset \}$ a $\sum_{i=1}^k p(H_i) = 1$ tak:

$$\sum_{i=1}^C p(H_k = i | D) = 1 \quad P(D) = \sum_{i=1}^C p(D | H_i) p(H_i)$$



1.10 Bayesovské učení

Bayesovské učení si klade dvě základní otázky:

- 1) Jaká hypotéza o modelovaném systému je s největší pravděpodobností platná?
- 2) Jaká je predikce nové instance na základě známých hypotéz?

Bayesovské učení, stejně jako podobné nástroje statistiky, slouží k poznávání a předvídání jevů v reálných podmínkách **neurčitosti**. Oproti klasickým statistickým metodám je však chápáno jako způsob myšlení, který umožňuje nerozporně využít teoretickou, experimentální a expertní znalost a také poskytuje nejen odhady neznámých veličin, ale i informaci o jejich přesnostech, a to i v konečných časech pozorování.(5)

Používají se tři druhy Bayesovských klasifikátorů **Brutální**, **Optimální** a **Naivní**. Brutální klasifikátor vybírá z množiny hypotéz hypotézy konzistentní s trénovacími

daty, které jsou si rovnocenné, přičemž ostatní hypotézy jsou vyloučeny. Při predikci jsou si i výstupy hypotéz rovnocenné. Optimální klasifikátor stanoví aposteriorní pravděpodobnost všech hypotéz. Při predikci jsou výstupy hypotéz váhovány aposteriorními pravděpodobnostmi. Naivní klasifikátor provádí klasifikaci na základě nejpravděpodobnější klasifikace za předpokladu vstupního vektoru X . (8)

1.11 Naivní Bayesův klasifikátor

Klasifikační metody používané při analýze dat jsou metody, které se zabývají hledáním struktury a vzájemných vazeb v objektech. Sledujeme určité objekty, které se více či méně navzájem odlišují. Cílem je zařadit buď některé z objektů, nebo všechny objekty do skupin, popř. uspořádat skupinu objektů do vnitřně sourodých tříd či shluků. (5)

Naivní bayesův klasifikátor je jednoduchý pravděpodobnostní klasifikátor založený na užití Bayesova teorému s velmi silným (naivním) předpokladem nezávislosti. Vystižnějším názvem pro tento zásadní pravděpodobnostní model by byl model nezávislého rysu. (8)

Jednoduše řečeno, Naivní Bayesův klasifikátor předpokládá, že přítomnost (či absence) specifického rysu určité třídy nesouvisí s přítomností (nebo absencí) jakéhokoliv jiného rysu. (8)

Pro přiblížení, nějaké uvažované ovoce může být považováno za jablko, jestliže je červené, kulaté a má řekněme 10 cm v obvodu. I přesto, že všechny tyto rysy závisí jeden na druhém či na existenci ostatních rysů, Naivní Bayesův klasifikátor uvažuje, že všechny tyto vlastnosti nezávisle přispívají k pravděpodobnosti, že nějaké uvažované ovoce je jablko. (8)

Díky precizní podstatě tohoto pravděpodobnostního modelu, lze Naivní Bayesův klasifikátor velmi efektivně využít při nastavování neuronových sítí či informačních systému pomocí tzv. supervised learning. V mnoha praktických aplikacích, se při odhadu parametrů pro Naivní Bayesův klasifikátor využívá tzv. metoda maximální pravděpodobnosti. Jinými slovy, s Naivním Bayesovým klasifikátorem lze pracovat i

pokud nevěříme v Bayesovskou statistiku a není potřeba využívat jakékoliv Bayesovy metody. (8)

I přes jejich prosté a zjevně velmi zjednodušené předpoklady, naivní Bayesovy klasifikátory si počínali velmi dobře v mnoha komplexních skutečných situacích.

V roce 2004 provedené analýzy Bayesova klasifikačního problému ukázali, že zde existují určité teoretické důvody pro domněle nesmyslnou účinnost naivních Bayesových klasifikátorů. (8)

I přesto, souhrnné porovnávání s dalšími klasifikačními metodami v roce 2006 ukázalo, že Bayesova klasifikace je překonána současnými metodami, jako jsou např. tzv. Gradient Boosting rozhodovací stromy a tzv. náhodné lesy (RF - random forests). Za zmínku stojí, že náhodné lesy jsou nejslibněji rozvíjející se klasifikační metodou. Ta je založena na rozhodovacím stromu, nicméně neobsahuje jeden rozhodovací strom, ale celou množinu rozhodovacích stromů, která rozhoduje o zařazení objektu do daných tříd (klasifikace). Náhodné lesy se principiálně nesnaží minimalizovat chybu jednoho stromu, ale celého lesa.

Výhodou naivního Bayesova klasifikátoru je, že vyžaduje pouze malé množství trénovacích dat pro odhad parametrů nutných ke klasifikaci. Protože nezávislé proměnné jsou předpokládány, musí být stanoveny pouze rozptyly proměnných pro každou třídu a ne celá kovarianční matice .

1.12 Naivní Bayesův pravděpodobnostní model

Pravděpodobnostní model pro klasifikaci je ve své podstatě podmíněným modelem.

$$p(C|F_1, \dots, F_N)$$

přes závislé proměnné třídy C s malým počtem výstupů nebo tříd, závisí na několika budoucích proměnných F_1 až F_n . Problém nastane, pokud je počet charakteristik n velký, nebo pokud charakteristika nabývá velkého množství hodnot. Potom je založení

modelu na pravděpodobnostních tabulkách neproveditelné. Je proto potřeba přeformulovat model pro lepší učenlivost. Použitím Bayesova vzorce dostaneme:

$$p(C | F_1, \dots, F_N) = \frac{p(C)p(F_1, \dots, F_N)}{p(F_1, \dots, F_N)}$$

V jednoduché řeči se předchozí vzorec dá přepsat jako:

$$\text{aposteriomi} = \frac{\text{apriorní} * \text{pravdepodobnost}}{\text{evidence}}$$

V praxi nás zajímá pouze čítec rovnice, pokud vycházíme z předpokladu, že jmenovatel je nezávislý na C a hodnoty F_i jsou dané. Potom je jmenovatel v podstatě konstantou. Čítec je pak ekvivalentní spojenému pravděpodobnostnímu modelu:

$$p(C | F_1, \dots, F_N)$$

který může být přepsán následovně, za použití opakované aplikace definice pro podmíněnou pravděpodobnost:

$$\begin{aligned} p(C | F_1, \dots, F_N) &= p(C)p(F_1, \dots, F_N | C) \\ &= p(C)p(F_1 | C)p(F_2, \dots, F_N | C, F_1) \\ &= p(C)p(F_1 | C)p(F_2 | C, F_1)p(F_3, \dots, F_N | C, F_1, F_2) \\ &= p(C)p(F_1 | C)p(F_2 | C, F_1)p(F_3 | C, F_1, F_2)p(F_4, \dots, F_N | C, F_1, F_2, F_3) \end{aligned}$$

2 Analýza problému a současná situace

Snaha o kvalitní predikci vývoje kapitálového trhu, ať již jako celku nebo jen jeho částí – akcií, provází trh od nepaměti. Vývoj tohoto de facto matematického odvětví zaznamenal značný pokrok co do sofistikovanosti zkoumaných řešení, nicméně úspěšnost je stále neuspokojující. Ba co víc, v dnešní době nejsme ani s to zodpovědět jednoduchou otázku, a to zda-li je predikce chování tak ohromného organismu, jakým kapitálový trh je, vůbec možná. Mnoho odborníků není jednotných ani v samotném popisu struktury kapitálového trhu. Například Edgar E. Peters se zabývá otázkou, zda-li jsou finanční trhy fraktální, což by ve své podstatě popíralo teorii efektivních trhů.

Vzhledem k rozporům v teorii finančních trhů jako takových, lze nalézt také celou řadu metod pro jejich predikci. Od základních metod jakými jsou technická nebo fundamentální analýza, přes regresi a klouzavé průměry, velké množství ukazatelů a indexů až k současnému trendu a tím je právě umělá inteligence. V této práci, jak již bylo zmíněno, jde zejména o aplikaci Bayesova Naivního klasifikátoru. Důvod, proč jsem se rozhodl využít tuto ne zcela běžnou metodu, je její intenzivní studium v posledních letech. Přestože bylo v roce 2006 prokázáno, že Bayesovská klasifikace je překonána současnými metodami Gradient Boosting a teorií náhodných lesů a přestože jsem se s prediktivním modelem založeným na bázi NBK pro finanční trhy nesetkal, rozhodl jsem se tuto, v poslední době velmi skloňovanou metodu použít.

3 Vlastní návrh řešení – implementace NBK

V této části práce se pokusím navrhnout Naivní Bayesův Klasifikátor v prostředí matematického programovacího jazyka MATLAB. Výběr tohoto prostředí není klíčový, zdrojový kód je možno napsat v jakémkoli programovacím jazyce, není dokonce nutná ani objektová orientovanost. V tomto případě je MATLAB zvolen díky snadné grafické reprezentaci výsledků, kterými jsou v našem případě grafy průběhů predikovaných a reálných hodnot akcie. V první části této kapitoly se pokusím nejprve vytvořit prediktivní algoritmus založený na NBK, otestovat ho na obecných (uměle vytvořených datových souborech) a v případě funkčnosti modelu se jej pokusím implementovat pro reálné historické hodnoty konkrétní akcie.

3.1 Programové prostředí MATLAB

Jak jsem zmínil v předchozím odstavci, pro návrh algoritmu použiji jazyk matematického programu MATLAB, který je ve své podstatě shodný s programovacím jazykem C a má širokou škálu nástrojů pro vizualizaci výstupních dat. V mém případě se jednalo o verzi MathWorks MATLAB 7.8.0 R2009a. MATLAB umožňuje 3 základní režimy práce :

Dialogový režim

- je přístupný v okně Command Window
- v tomto režimu je možno MATLAB používat jako více méně inteligentní kalkulačku s funkcemi
- napsané výrazy/příkazy se po odeslání (stiskem klávesy ENTER) ihned vyhodnocují/vykonávají
- umožňuje práci s proměnnými pracovního prostředí (Workspace)

Programový režim

- slouží pro editaci M-souborů (skriptů nebo uživatelských funkcí)
- je spjat se samostatným oknem Meditoru (= editor M-souborů, kde je barevně zvýrazňována syntaxe příkazů)
- obsahuje debugger, který umožňuje hledat chyby ve funkcích/skriptech

Grafický režim

- slouží pro vizualizaci výsledků
- otevírá samostatná okna s názvem Figure

3.2 NBK - algoritmus

3.2.1 Prerekvizity z teoretické části

Jak již bylo zmíněno v teoretické části, Bayesovská klasifikace je principem blízká lidskému logickému uvažování. Umožňuje uvést pozdější zkušenosti do souladu s původními předpoklady.

Naivní Bayesovský klasifikátor vychází z Bayesovy podmíněné pravděpodobnosti. Je označován jako naivní, protože oproti ostatním Bayesovským klasifikacím (brutální, optimální) přináší značné zjednodušení výpočtu, přestože dosahuje velmi zajímavých výsledků.

Klasifikace na základě nejpravděpodobnější klasifikace g_{MAP} za předpokladu vstupního vektoru (x_1, \dots, x_n)

$$g_{MAP} = \arg \max_{j=1..C} P(g_j | x_1, \dots, x_n)$$

C – počet tříd

N – počet atributů

Tuto klasifikaci lze upravit následujícím způsobem:

$$g_{MAP} = \arg \max_{j=1 \dots C} \frac{P(x_1, \dots, x_n | g_j) \cdot P(g_j)}{P(x_1, \dots, x_n)} = \arg \max_{j=1 \dots C} P(x_1, \dots, x_n | g_j) \cdot P(g_j)$$

$P(g_j)$ – určí se na základě četnosti výskytu v trénovacích datech

$P(x_1, \dots, x_n | g_j)$ – určí se na základě zjednodušeného předpokladu, že hodnoty vstupních veličin jsou na sobě podmíněně nezávislé.

Pro výpočet sdružené podmíněné pravděpodobnosti za předpokladu nezávislosti platí:

$$g_{MAP} = P(x_1, \dots, x_n | g_j) = \prod_{i=1 \dots n} P(x_i | g_j)$$

Konečný vztah pro výslednou klasifikaci je pak dán vztahem:

$$g_{MAP} = \arg \max_{j=1 \dots C} P(g_j) \cdot \prod_{i=1 \dots n} P(x_i | g_j)$$

Problém může nastat v situaci, kdy konkrétní hodnota atributu pro danou třídu nikdy nenastala. Celý součin je pak kvůli jednomu členu roven 0, což není vhodné. To se dá řešit upravením výpočtu podmíněné pravděpodobnosti, nebo tvrdým ošetřením. V této diplomové práci je tento jev ošetřen tím způsobem, že je nulovým členům přiřazena hodnota blízká nule (konkrétně 0,001).

3.3 NBK implementace

3.3.1 Data pro obecnou implementaci

Prvním krokem je zvolení datového souboru pro otestování funkčnosti predikce na základě pravděpodobnostní klasifikace tříd vstupů. Datovým souborem je tabulka formátu MS Excel, kde každý řádek reprezentuje jednu sadu vstupních parametrů a jeden výstup z 5 možných stavů daného dne (viz. obr. 6).

Uvažoval jsem 4 vstupní parametry:

- den v týdnu – nabývá hodnot $\langle 1;7 \rangle$, což odpovídá dnům pondělí až neděle
- minulý vývoj akcie – nabývá hodnot $\langle -1;1 \rangle$;
-1 .. akcie klesala,
0 .. akcie stagnovala,
1 .. akcie rostla
- předminulý vývoj akcie – nabývá hodnot $\langle -1,1 \rangle$;
-1 .. akcie klesala,
0 .. akcie stagnovala,
1 .. akcie rostla
- expertní odhad akcie – nabývá hodnot -1,0,1

Druhou částí řádku datového souboru je 5 výstupních stavů $m1 - m5$, přičemž nastat může pouze jeden z nich:

- $m1$ – akcie klesá hodně
- $m2$ – akcie klesá málo
- $m3$ – akcie stagnuje
- $m4$ – akcie roste málo
- $m5$ – akcie roste hodně

Ve zdrojovém kódu obecné implementace jsou pak tyto výstupní stavy reprezentovány hodnotami -3, -1, 0, 1, 3, které se v závislosti na odhadu přičítají nebo odečítají z hodnoty pomyslné akcie. V této fázi obsahuje datový soubor 9 týdnů fiktivního vývoje akcie, tedy 63 sad hodnot vstupních parametrů a jednoho odpovídajícího výstupu pro

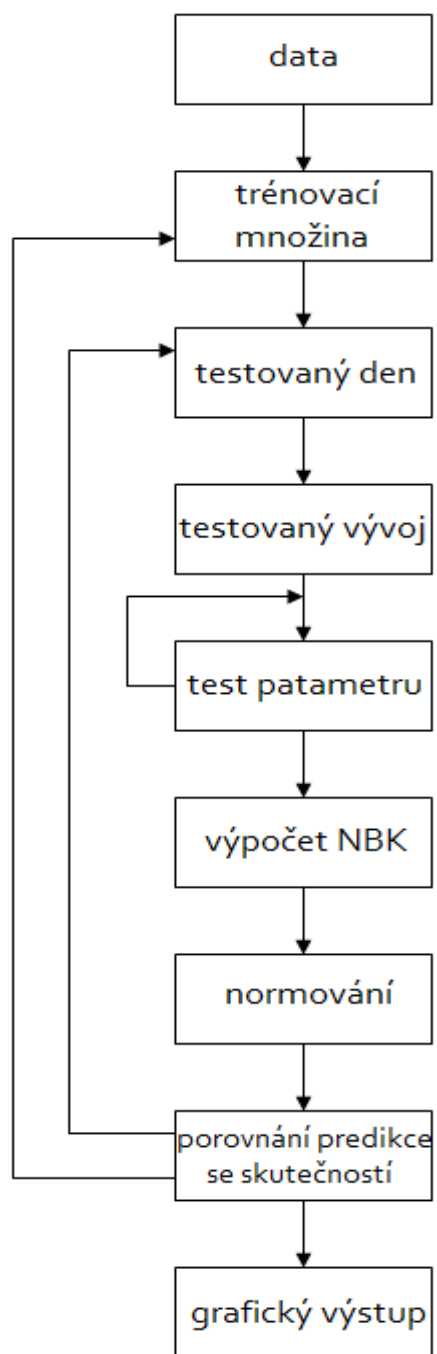
každý den. Pro snazší pochopení datového vstupu algoritmu popíši první řádek datového souboru z obr. 6 :

| | A | B | C | D | E | F | G | H | I |
|----|---|----|----|----|---|---|---|---|---|
| 1 | 1 | -1 | -1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 2 | 2 | -1 | -1 | -1 | 1 | 0 | 0 | 0 | 0 |
| 3 | 3 | -1 | -1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 4 | -1 | -1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 5 | 5 | -1 | -1 | 3 | 0 | 0 | 0 | 1 | 0 |
| 6 | 6 | -1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 7 | 7 | -1 | -1 | -1 | 0 | 0 | 1 | 0 | 0 |
| 8 | 1 | 1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 9 | 2 | 1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 10 | 3 | 1 | -1 | 3 | 0 | 0 | 0 | 0 | 1 |
| 11 | 4 | 1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | 5 | 1 | -1 | -1 | 0 | 1 | 0 | 0 | 0 |
| 13 | 6 | 1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 14 | 7 | 1 | -1 | 0 | 0 | 0 | 1 | 0 | 0 |

Obrázek 6: Datový soubor pro NBK

První den v týdnu (sloupec A, hodnota 1) platilo, že minulý vývoj akcie byl klesající (sloupec B, hodnota -1) a stejně tak předminulý vývoj akcie byl klesající (sloupec C, hodnota -1). Expertní odhad chování akcie byl stagnace (sloupec D, hodnota 0). Akcie však zareagovala při daných faktorech mírným poklesem (sloupec F, hodnota 1) – modrá část tedy tvoří vstupy, následuje jeden červený výstup pro každý den.

3.3.2 NBK – vývojový diagram algoritmu



Obrázek 7: Vývojový diagram NBK

3.3.3 NBK – popis algoritmu

Nejprve je z importovaných dat vybrána trénovací množina. Její naplnění je náhodné a pro každé spuštění programu je tedy jiná trénovací množina. Její velikost je stanovena na tři týdny, což v praxi znamená, že je vybráno třikrát po každém dni v týdnu. Algoritmus totiž nebere dny po sobě jdoucí, ale seskupuje stejné dny k sobě.

Následuje náhodný výběr situace, která v daný den nastává. Tedy postupně opět dny v týdnu s příslušnými vstupními parametry.

Poté se sekvenčně testují výstupní stavy chování akcie – program testuje jednotlivé vstupní parametry a počítá pravděpodobnost s jakou daná změna nastane či nenastane. Tyto pravděpodobnosti jsou normovány a je možno do této části kódu zahrnout například rozhodování o výhodnosti či nevýhodnosti investice.

Následně se porovnává skutečnost s vypočteným odhadem. V případě, že odhad odpovídá skutečnosti, inkrementuje se ukazatel úspěšnosti. V případě, že se odhad liší od skutečnosti, je ukazatel úspěšnosti dekrementován.

Pro potřeby grafického výstupu jsou hodnoty odhadnuté i skutečné v proměnných převedeny na hodnoty odpovídající pohybu akcie. Po všech potřebných přiřazeních je výsledek vykreslen ve třech základních grafech, které dokumentují výsledky dosažené algoritmem.

3.3.4 NBK – vlastní algoritmus

Tato kapitola je věnována popisu zdrojového kódu vlastního prediktivního algoritmu založeného na NBK. Pokusím se přehledně rozdělit kód do dílčích pasáží s vlastním popisem pro jeho snazší pochopení. Řádky uvozené znakem % jsou komentáři ve vlastním zdrojovém kódu a ponechám je beze změny. Plní doplňující funkci psaného popisu.

Po deklaraci funkce **NBK ()** následuje popis proměnných odpovídajících jednotlivým parametrům, tak jak jsem je definoval v kapitole 3.3.1 Data pro obecnou implementaci.


```

function NBK (data);

% data(i) = [den v tydnu , minuly vyvoj , predminuly vyvoj , expertni
ocekavani , aktualni stav]
% data(i) = [a , b , c , d, m1 , m2, m3, m4 , m5]
% m(r) - konkretni zmena akcie v zavislosti na parametrech a-d
% a - den v tydnu a = <1,7> (1-po , 2-ut , 3-st , 4-ct , 5-pa , 6-so ,
7-ne)
% b - minuly vyvoj b = <-1,1>
% c - predminuly vyvoj c = <-1,1>
% e - expertni ocekavani k = <-1;0;3>

```

V této části kódu probíhá trénování. V podstatě se jedná jen o naplnění trénovací množiny **TREN()** daty. První cyklus **for i = 1:7** zařídí, že program vybere 7 po sobě jdoucích dní. Druhý, vnořený cyklus **for j = 1:5** definuje délku trénovacího procesu v týdnech a lze libovolně měnit.

```

close all;

index = 1;
for i = 1:7
    % trenovani trva pet tydny
    for j = 1:5
        in = i+(7* round(8*rand));
        TREN(index,:) = data(in,:);
        index = index+1;
    end
end

```

Zbývající kód až do části grafického zpracování je proces učení. Vzhledem k rozsáhlosti procesu budou dílčí operace komentovány. Nabytá zkušenost se přidává k trénovací množině. Délku procesu učení v týdnech definuje první cyklus **for la = 1:5**. Obecně však neplatí, že delším učením dosáhne model lepších výsledků. Po určité době dochází k jevu, který se nazývá přeučení a je popsán v kapitole Strojové učení.

```

% in_st , c_nbk , c_nah - promenne pro graficke zpracovani vylsedku

in_st = 1;
c_nbk = zeros(63,5);
c_nah = zeros(63,5);
for la = 1:5
    % nahodny vyber situaci jednoho tydne postupne podle dni
    for i = 1:7

        in = i+(8* round(6*rand));
        TEST = data(in,:);
    end
end

```

Po fázi načtení náhodně vybraných dat z datového souboru do matice **TEST()** se testují výstupní stavy v trénovací množině **TREN()**. Pokud nastane shoda, inkrementuje se pro daný výstup pravděpodobnost **Da**. Pokud by žádný z výstupů nenabýval hodnoty 1, tzn. že by pro danou sadu vstupních parametrů nenastala žádná definovaná situace (pouze teoreticky, v datovém souboru musí alespoň jeden ze stavů nastat), došlo by při výpočtu pravděpodobnosti kvůli jednomu nulovému členu k vynulování celého součinu. Proto následuje z bezpečnostních důvodů (chyba v datovém souboru, popřípadě nám neznámý výstup pro danou sadu vstupů) ošetření krajních případů a k pravděpodobnosti **Da** je připočtena velmi malá hodnota, blízká nule, tak jak je to popsáno v kapitole 3.2.1 Prerekvizity z teoretické části.

```

% vypocet pro jednotlivé zmeny vyvoje
% r - konkrétní situace
for r = 1:5

    % NASTANE ZMENA
    % -----
    [x,y] = size(TREN);
    Da = 0;
    for j = 1:x
        if TREN(j,4+r) == 1
            Da = Da+1;
        end
    end
    %podminka osetreni krajnich pripadu
    if Da == 0
        Da = 0.001;
    end
    PDa = Da/x;
end

```

V další části testování se načítají postupně vstupní parametry příslušného řádku datového souboru a ukládají se do proměnné `a`. Následně se porovnává vstupní parametr z trénovací množiny s odpovídajícím parametrem z množiny testovací (`if TREN(j,k) == a(k)`) přičemž pokud jsou shodné a byla v jejich případě zaznamenána změna na dané pozici ve výstupních stavech (`& TREN(j,4+r) == 1`), inkrementuje se její pravděpodobnost `Pka(k)`. Následuje opět ošetření krajních případů jako v předchozí části kódu.

```
% ostatni vypocty parametru
% k - vybrany parametr - sloupec matice data
% p(k) - pocet pro který splnuje
% Pka(k) - pravdepodobnost tohoto parametru

for k = 1:4
    a(k) = TEST(:,k);
    p(k) = 0;
    for j = 1:x
        if TREN(j,k) == a(k) & TREN(j,4+r) == 1
            p(k) = p(k)+1;
        end
    end
    %podminka osetreni krajnich pripadu
    if p(k) == 0
        p(k) = 0.001;
    end
    Pka(k) = p(k)/Da;
end
```

Výpočet pro pravděpodobnost, že za daných parametrů situace nenastane, probíhá analogicky s tím rozdílem, že se netestuje jestli změna při daných parametrech nastala, ale zda nenastala, tzn. zda byla různá od 1 (`if TREN(j,4+r) ~= 1`).

```

% NENASTANE ZMENA
[x,y] = size(TREN);
Dn = 0;
for j = 1:x
    if TREN(j,4+r) ~= 1
        Dn = Dn+1;
    end
end
%podminka osetreni krajnich pripadu
if Dn == 0
    Dn = 0.001;
end
PDn = Dn/x;

% ostatni vypocty parametru
% k - vybrany parametr - sloupec matice data
% Pkn(k) - pravdepodobnost tohoto parametru

for k = 1:4
    a(k) = TEST(:,k);
    p(k) = 0;
    for j = 1:x
        if TREN(j,k) == a(k) & TREN(j,4+r) ~= 1
            p(k) = p(k)+1;
        end
    end
    %podminka osetreni krajnich pripadu
    if p(k) == 0
        p(k) = 0.001;
    end
    Pkn(k) = p(k)/Dn;
end

```

Vlastní výpočet pravděpodobností je pak prováděn podle Bayesova vzorce podmíněné pravděpodobnosti, tak jak je uveden v předešlých kapitolách teoretické části. V tomto případě se pravděpodobnost daného výstupního stavu skládá z pravděpodobností každého parametru, pro které nastává změna **Pka(1) – Pka(4)** a dále z pravděpodobnosti změny daného stavu **PDa**. Tyto všechny pravděpodobnosti jsou pak normovány podle základního Bayesova vzorce **A(r) = Pa/(Pa+Pn)**. Analogicky je pak proveden výpočet pravděpodobností, že za daných podmínek změna nenastane.

```

% vypocet pravdepodobnosti ze ne/nastane konkretni zmena
% r - konkretni zmena
% A - nastane
% N - nenastane
% -----
Pa(r) = PDa*Pka(1)*Pka(2)*Pka(3)*Pka(4);
Pn(r) = PDn*Pkn(1)*Pkn(2)*Pkn(3)*Pka(4);
A(r) = Pa / (Pa+Pn);
N(r) = Pn / (Pa+Pn);

```

Jednoduchým testem zjistíme, zda-li byla algoritmem vypočtená predikce správná či nikoli. Do proměnné **odhad** přiřadíme hodnotu 0 nebo 1 podle toho, která z pravděpodobností má vyšší váhu. Tuto proměnnou následně porovnáme se skutečností, která je načtena z datového souboru. Na základě porovnání poté v případě shody inkrementujeme ukazatel úspěšnosti **c_nbk**, v případě neshody dekrementujeme. Získáme tím po proběhnutí programu grafický přehled o stupni vývoje úspěšnosti predikce.

```

% urceni spravnosti odhadu
skutečnost = TEST(:,4+r);

if A(r) < N(r)
    odhad = 0;
elseif A(r) > N(r)
    odhad = 1;
else
    disp('chyba');
end

% ukazatel uspesnosti
pltx(r,in_st) = in_st;
if skutečnost == odhad
    c_nbk(r) = c_nbk(r)+1;
    %disp('');
end
if skutečnost < odhad
    c_nbk(r) = c_nbk(r)-1;
    %disp('');
end
if skutečnost > odhad
    c_nbk(r) = c_nbk(r)-1;
    %disp('');
end
pltz(r,in_st) = c_nbk(r);
end

```

Předposlední částí programu je naplnění proměnných pro grafické zpracování. Nejdříve zjistíme pozici maximální hodnoty v proměnné A, která reprezentuje vlastní odhad. To znamená, že v podstatě hledáme pozici čísla 1. Tato hodnota nám určí význam odhadu ve smyslu pohybu akcie – pozice 1,2,3,4,5 odpovídají změnám hodnoty akcie -3,-1,0,+1,+3. Přepínačem poté zajistíme správné přiřazení hodnot pro přičtení k poslednímu stavu akcie. Následně je přidána kompletní zkušenost, kterou představují skutečná data, k trénovací množině.

```
[def,i] = max(A);

% naplneni dat odhadnutych
switch i
    case 1
        ii = -3;
    case 2
        ii = -1;
    case 3
        ii = 0;
    case 4
        ii = 1;
    case 5
        ii = 3;
end

pltn(1, in_st) = ii;

if in_st == 1
    plty(1,in_st) = ii;
else
    plty(1,in_st) = plty(1,in_st-1) + ii;
end

[def,j] = max(data(in,5:9));

switch j
    case 1
        jj = -3;
    case 2
        jj = -1;
    case 3
        jj = 0;
    case 4
        jj = 1;
    case 5
        jj = 3;
end

pltn(2, in_st) = jj;

% naplneni dat realnych
if in_st == 1
```

```

        plty(2,in_st) = jj;
    else
        plty(2,in_st) = plty(2,in_st-1) + jj;
    end

    pltx(1,in_st) = in_st;
    % pridani kompletne zkusenosti k trenovacim datum
    in_st = in_st+1;
    TREN(index,:) = data(in,:);
    index = index+1;
end
end

```

Poslední částí zdrojového kódu je vykreslení grafů pro posouzení úspěšnosti.

```

% graficke vyjadreni odhadu vs skutečnosti v case
figure();
plot(pltx(1,:),plty);
title('Graf vyvoje akcie v case');
legend('bayes','real');
xlabel('pocet dnu \rightarrow');
ylabel('vyvoj akcie \rightarrow');
%graficke vyjadreni odhadu prirustku vs skutečnosti v case
figure();
plot(pltx(1,:),pltn);
title('Graf vyvoje prirustku v case');
legend('bayes','real');
xlabel('pocet dnu \rightarrow');
ylabel('vyvoj akcie \rightarrow');
% graficke vyjadreni uspesnosti UCENI pro jednotlivé vystupy
figure();
plot(pltx(1,:),pltz);
title('Graf uspesnosti odhadu pro jednotlivé stavy');
legend('roste  hodne','roste  malo','stagnace','klesa  hodne','klesa  malo');
xlabel('pocet dnu \rightarrow');
ylabel('uspesnost uceni \rightarrow');

```

3.4 Testování na obecných datech

V této kapitole bude předvedeno několik pokusů s algoritmem NBK na obecné datové sadě. Jde tedy v podstatě o jakýsi test funkčnosti, který však nebude prováděn na reálných datech, protože je nejprve nutné ověřit funkčnost algoritmu, zda je možno vůbec nalézt závislosti pro klasifikaci.

Testovací datový soubor reprezentuje 9 týdnů, tedy 63 dní vývoje pomyslné akcie. Obsahuje, jak již bylo zmíněno, 4 vstupní parametry ovlivňující vývoj akcie, který je uveden jako 1 z 5 možných výstupů.

Výstupem pro tuto kapitolu jsou 3 grafy:

GRAF 1

Tento graf znázorňuje predikci a skutečný vývoj akcie v čase. Hodnoty skutečného vývoje jsou čerpány z datového souboru data.xls. Počátek vykreslování je v nule, jelikož není k dispozici žádná výchozí hodnota akcie. Ostatně to pro predikci ani není nutné, jelikož nás zajímá chování v čase a počáteční hodnota akcie není podstatná.

Skutečný vývoj je v grafu vykreslen zeleným perem, v legendě uveden jako REAL.

Predikovaný vývoj akcie je v grafu znázorněn modře a v legendě je uveden popis BAYES. Modrá křivka je tedy předmětem zkoumání.

GRAF 2

V tomto grafu je znázorněn průběh skutečných a předikovaných přírůstků akcie v čase. Je podstatný z prostého důvodu. Pokud započne predikce špatným odhadem, což je z logiky algoritmu běžný jev (algoritmus se učí, proto správný odhad již v počátku je dilem spíše štěstí než predikce), může se nastřádat rozdíl mezi skutečnou hodnotou a predikcí. V praxi to pak znamená, že pokud se odhad nebude shodovat se skutečnou hodnotou třeba jen v prvních dvou krocích, počáteční rozdíl naroste o 12 a Graf 1 se tak bude jevit jako lichý. Nicméně v tomto případě by byl opomenut fakt, že další predikce se může přibližovat skutečným hodnotám, ale v důsledku rozdělení křivek v počátku to nebude patrné. Graf 2 je popsán stejně jako předchozí, proto zelená křivka reprezentuje

hodnoty skutečných přírůstků hodnot akcie a modrá pak hodnoty predikovaných přírůstků.

GRAF 3

Posledním grafem je ukazatel úspěšnosti odhadů pro jednotlivé parametry. Výstupem ovšem není procentuelní úspěšnost, nýbrž tzv. počet zásahů, tedy úspěšných odhadů. Odhad je navíc při neúspěchu penalizován, což lze využít při váhování pravděpodobností hypotéz. Nicméně na vysvětlenou, pokud při desátém dni lze z grafu ukazatelů úspěšnosti u některého z nich odečíst hodnotu například 8, znamená to, že počet úspěšných odhadů je 8 z celkových 10, a tím tedy algoritmus dosáhl u tohoto ukazatele 80% úspěšnost.

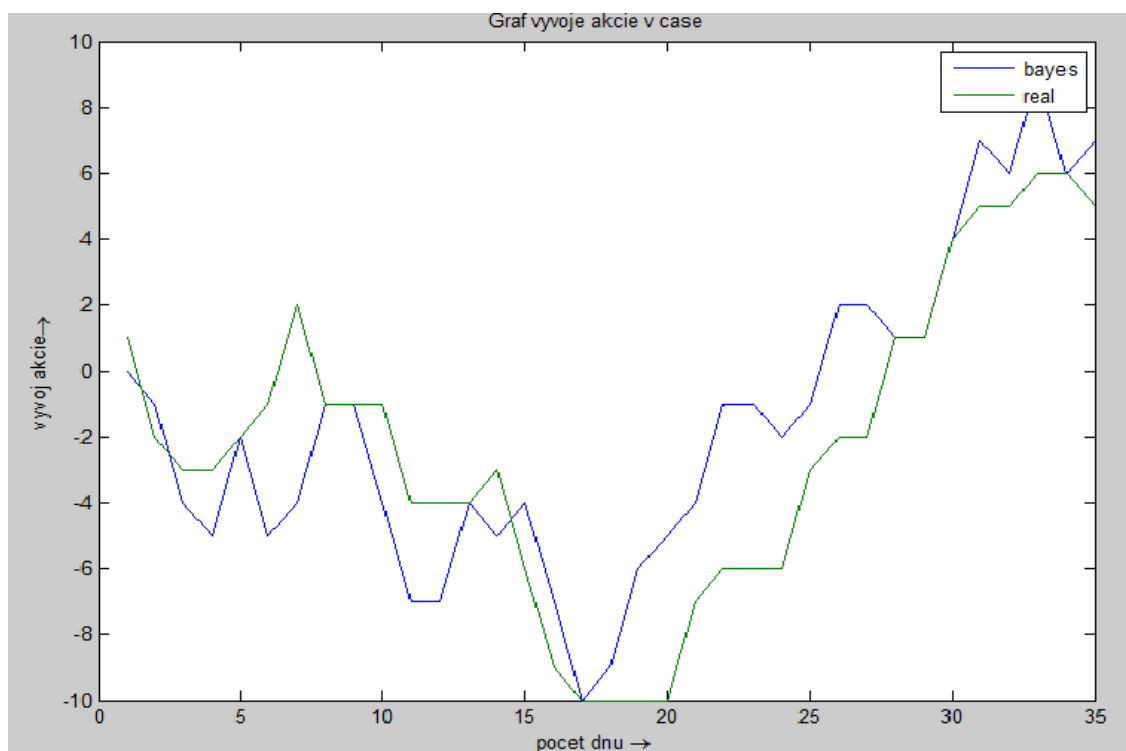
3.5 TEST 1 – 3x5

| | |
|---------------|---------|
| Délka pokusu | 5 týdnů |
| Trénování | 3 týdny |
| Počet iterací | 175 |

První sada testů je tedy determinována těmito parametry: délka trénování 3 týdny, což znamená že se vytvoří trénovací množina o 21 prvcích (dnech). Následný proces učení a predikce trvá 5 týdnů. Nejde o 5 týdenní predikci vývoje akcie dopředu ale predikuje se dopředu jeden krok. Nicméně snahou této práce je predikovat celý průběh vývoje akcie. Vlastní odhad vývoje ceny akcie je tedy porovnáván z hodnot naměřených v 35 dnech.

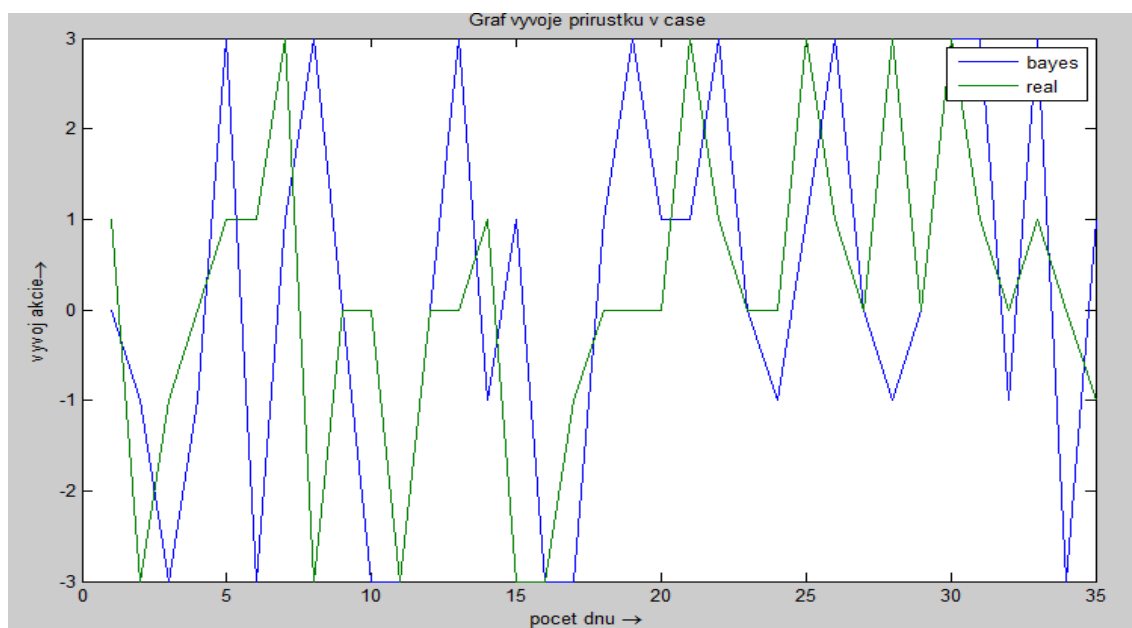
3.5.1 TEST 1.1

Výstupem prvního testu predikce je graf na obr. 8. V tomto prvním případě je vidět vcelku zajímavá úspěšnost algoritmu.



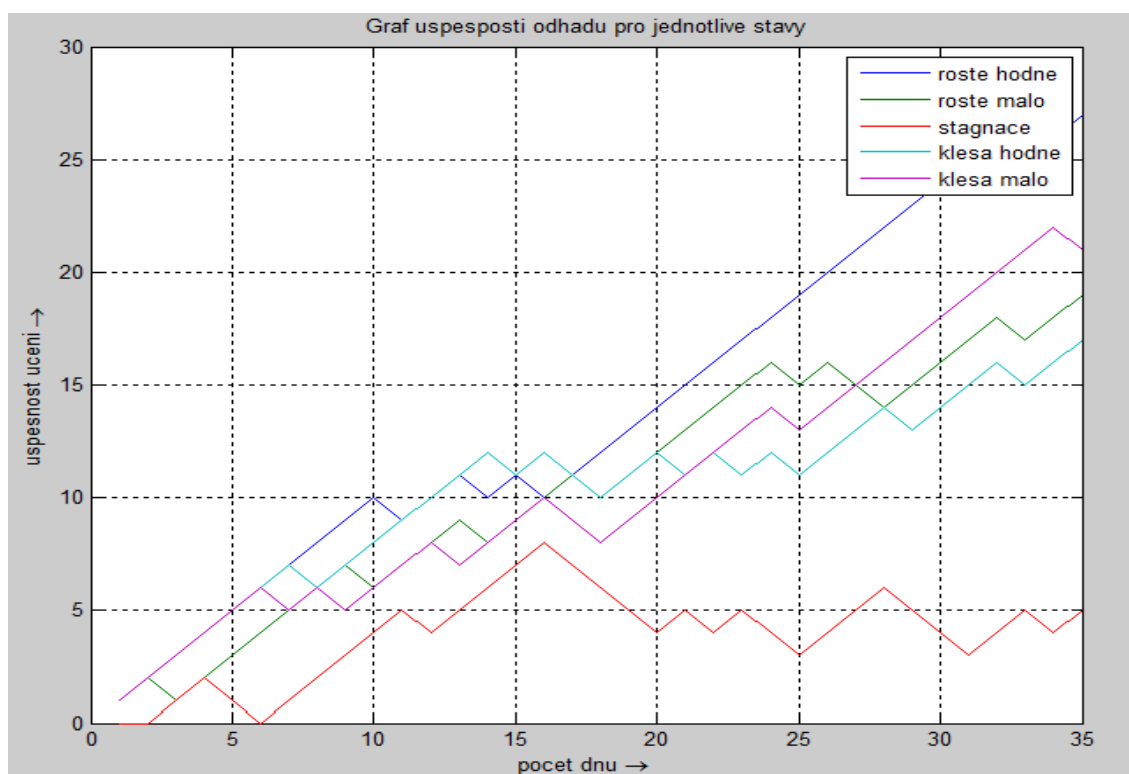
Obrázek 8: TEST 1.1 - vývoj akcie v čase - 3x5

Graf na obrázku 9 vyjadřuje shodu v absolutních přírůstcích predikce a skutečného stavu. Je vidět určitou oscilaci predikované hodnoty (modrá) kolem hodnot skutečných (zelená).



Obrázek 9: TEST 1.1 – vývoj přírůstků v čase – 3x5

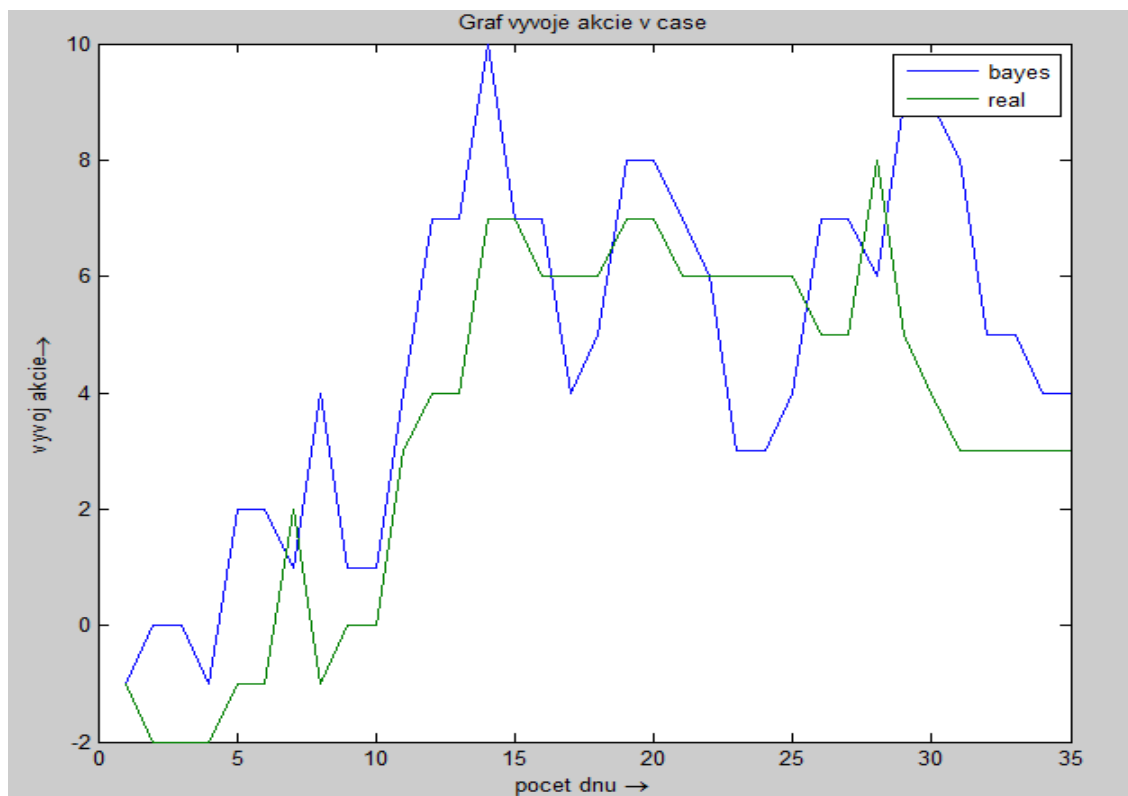
Graf na obrázku 10 znázorňuje ukazatel úspěšnosti pro jednotlivé výstupy. Tato úspěšnost je však v rámci normalizace pravděpodobností penalizována, a to tak, že je ukazatel při špatném odhadu dekrementován. To je patrné z klesajících částí grafu a pro grafickou interpretaci je to možná nešťastné. De facto to snižuje skutečnou úspěšnost, jelikož do proměnné ukazatelů každého parametru se zanáší v podstatě počet zásahů (správných odhadů), tak dekrementování této proměnné má za následek faktické snížení prezentované úspěšnosti modelu v grafu. Proto v dalších testech vytvořím nový graf, který bude čerpat data z další proměnné za stejných podmínek, nebude však dekrementovat ukazatel úspěšnosti. To umožní určení procentuelní úspěšnosti odhadu pro každý parametr.



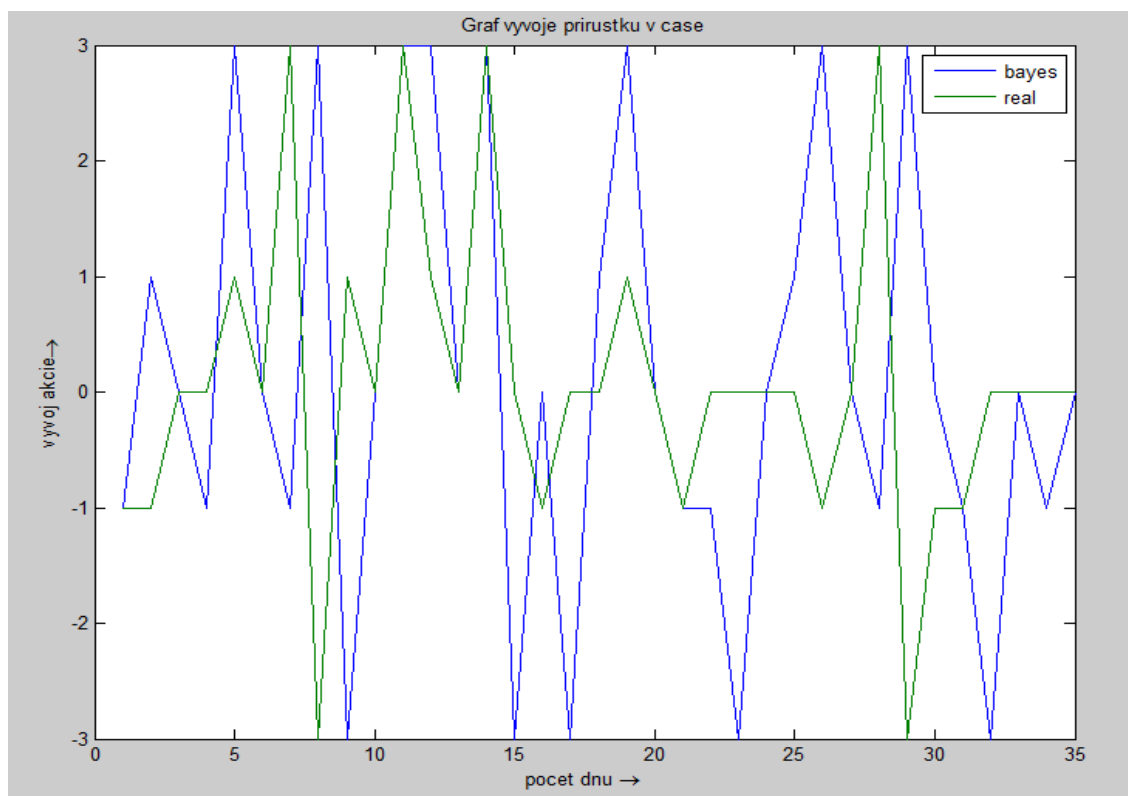
Obrázek 10: TEST 1.1 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5

3.5.2 TEST 1.2

Druhý test s identickými parametry. Vzhledem k tomu, že trénovací i testovací data jsou vybírána náhodně, je pro každé spuštění programu jiný výstup. Druhý pokus dopadl opět relativně dobře. Počáteční špatný odhad v prvních dvou dnech sice způsobil, že další špatný odhad v následujících dvou dnech přiblížil odhadnutou hodnotu akcie té reálné (na počátku se predikce a skutečnost vzdálila, avšak vlivem další chybné predikce se křivky opět přiblížily). Nicméně v dalších dnech až na drobné výkyvy osciluje predikovaný vývoj akcie kolem hodnoty reálné. Při tomto počtu kroků, tedy při 35 dnech zatím nedochází v tomto ani v předchozím pokusu k jevu přeučení.



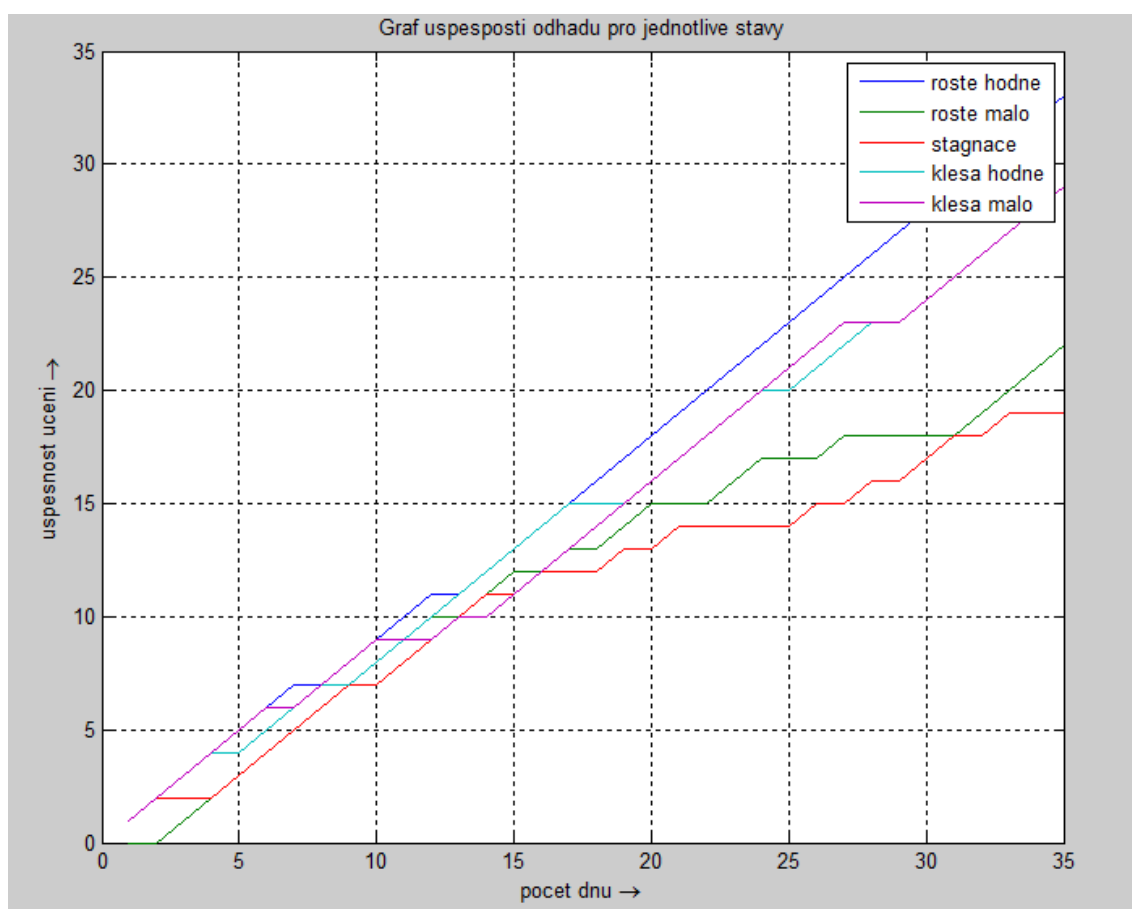
Obrázek 11: TEST 1.2 – vývoj akcie v čase – 3x5



Obrázek 12: TEST 1.2 - vývoj přírůstků v čase - 3x5

Graf ukazatelů úspěšnosti na obrázku 13 dovoluje odečíst hodnoty pro výpočet úspěšnosti jednotlivých ukazatelů:

- roste hodně 85 % úspěšnost
- roste málo 80 % úspěšnost
- stagnace 73 % úspěšnost
- klesá hodně 85 % úspěšnost
- klesá málo 73 % úspěšnost

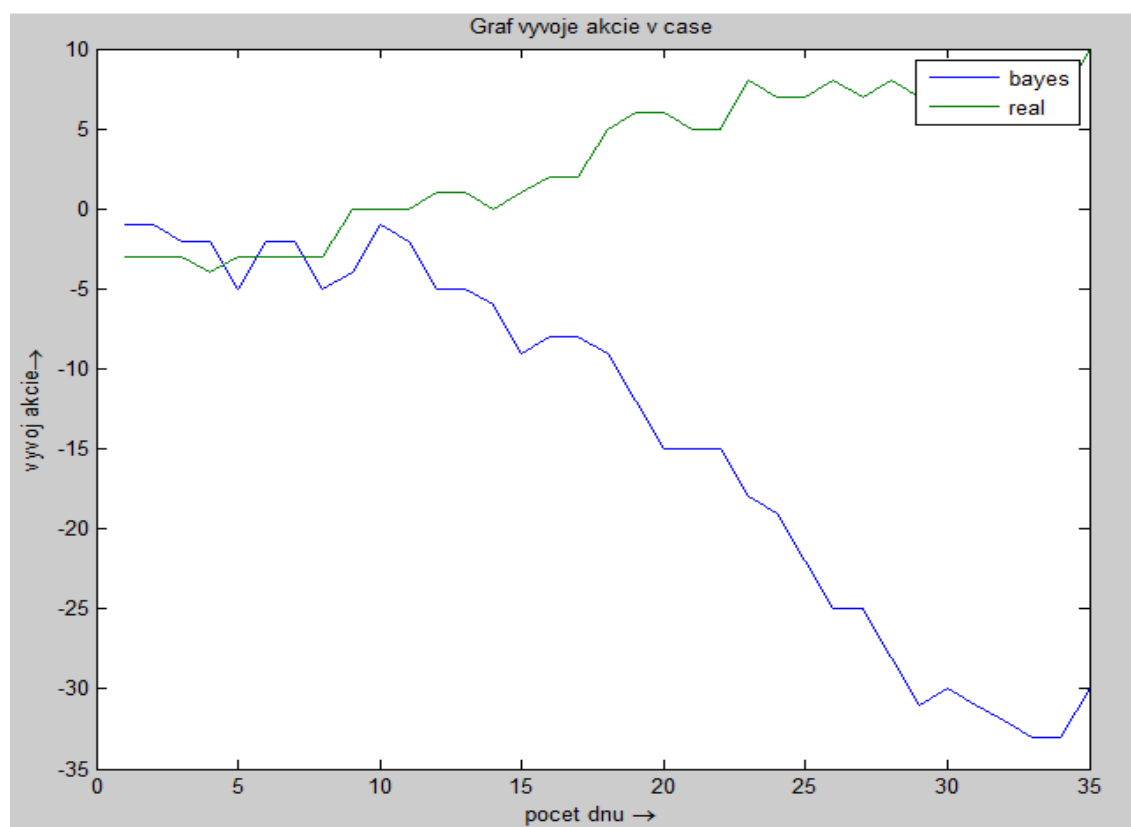


Obrázek 13: TEST 1.2 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5

3.5.3 TEST 1.3

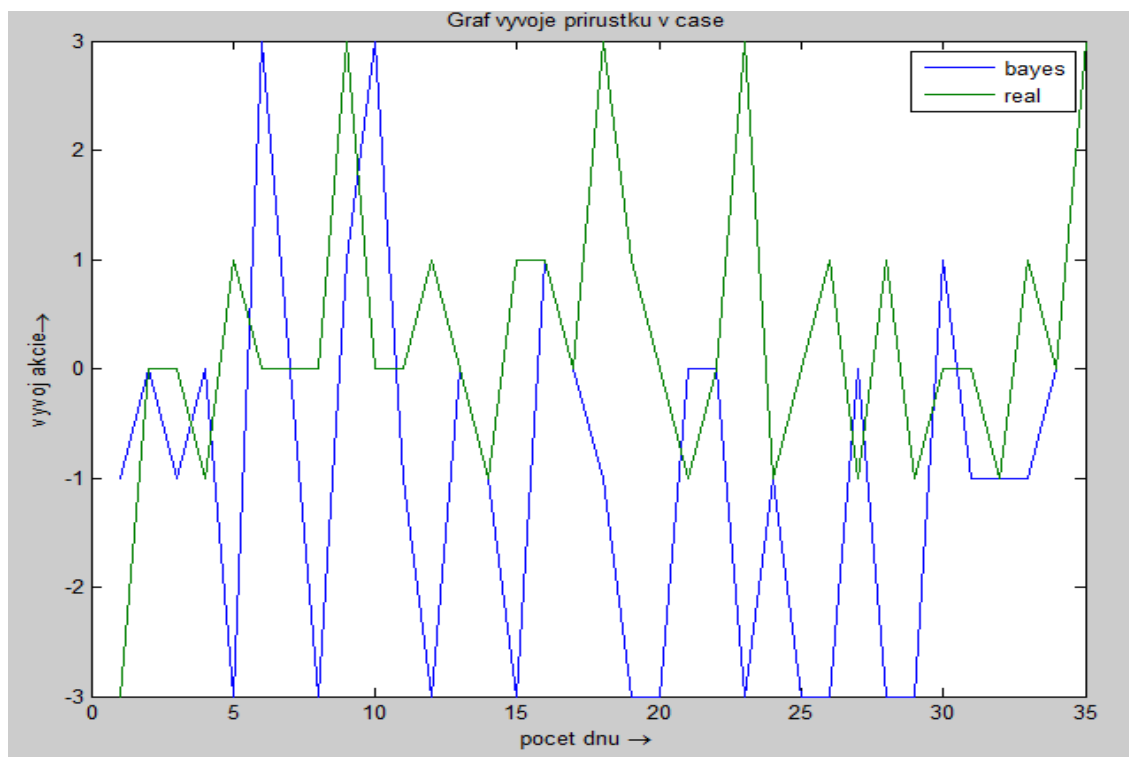
Tento test reprezentuje případ, kdy vyhodnocené pravděpodobnosti vstupů nemají žádný zásadní vliv na výstup. Jinými slovy, že pozorovaný výstup je na daných vstupních parametrech nezávislý. Je to pravděpodobně dáno rozložením fiktivních dat, nicméně akciový trh je plný stochastických jevů, a proto je predikce na omezeném počtu vstupů vždy úspěšná jen za určitých podmínek.

Není totiž možné zkoumat dostatečné množství vstupů ovlivňující výstup, protože je nikdy nebudeme znát všechny a co víc, mnoho z nich je ovlivněno například psychikou samotných lidí, a proto je nelze deterministicky definovat. Navíc obrovské množství faktorů, které mohou ovlivnit pohyb akcie, vyvolá zákonitě předčasné přeučení modelu, protože ten bude muset z tak velkého množství vyextrahovat závislosti, které ve skutečnosti závislostmi vůbec nejsou a budou tvořit pouze šum, který negativně ovlivní úspěšnost predikce.



Obrázek 14: TEST 1.3 - vývoj akcie v čase - 3x5

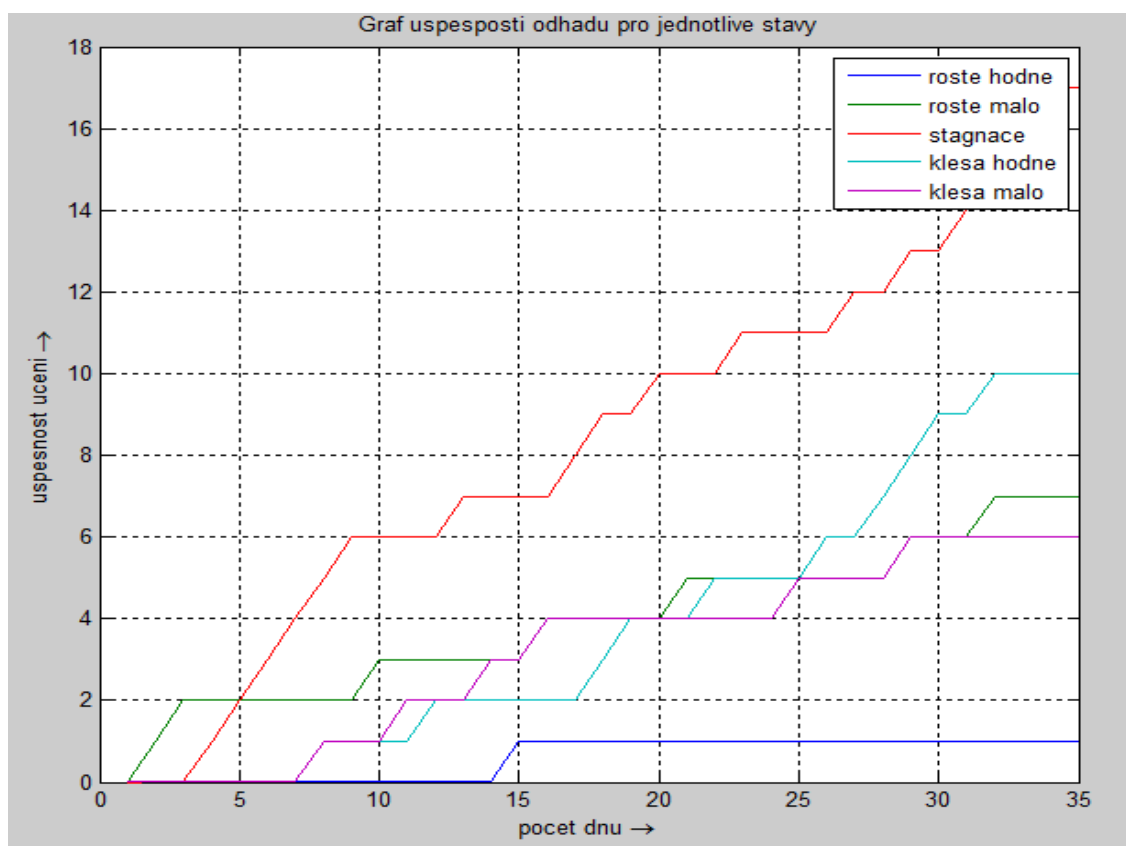
Na grafu přírůstků je patrné, že predikce je doslova převrácená naruby. Jak jsem již zmínil, je to způsobeno pravděpodobně velmi malou až žádnou závislostí výstupu na vstupních parametrech, a proto je predikce dalšího vývoje chybná.



Obrázek 15: TEST 1.3 - vývoj přírůstků v čase - 3x5

Hodnoty ukazatelů úspěšnosti dle grafu

- roste hodně 4 % úspěšnost
- roste málo 20 % úspěšnost
- stagnace 40 % úspěšnost
- klesá hodně 20 % úspěšnost
- klesá málo 20 % úspěšnost



Obrázek 16: TEST 1.3 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5

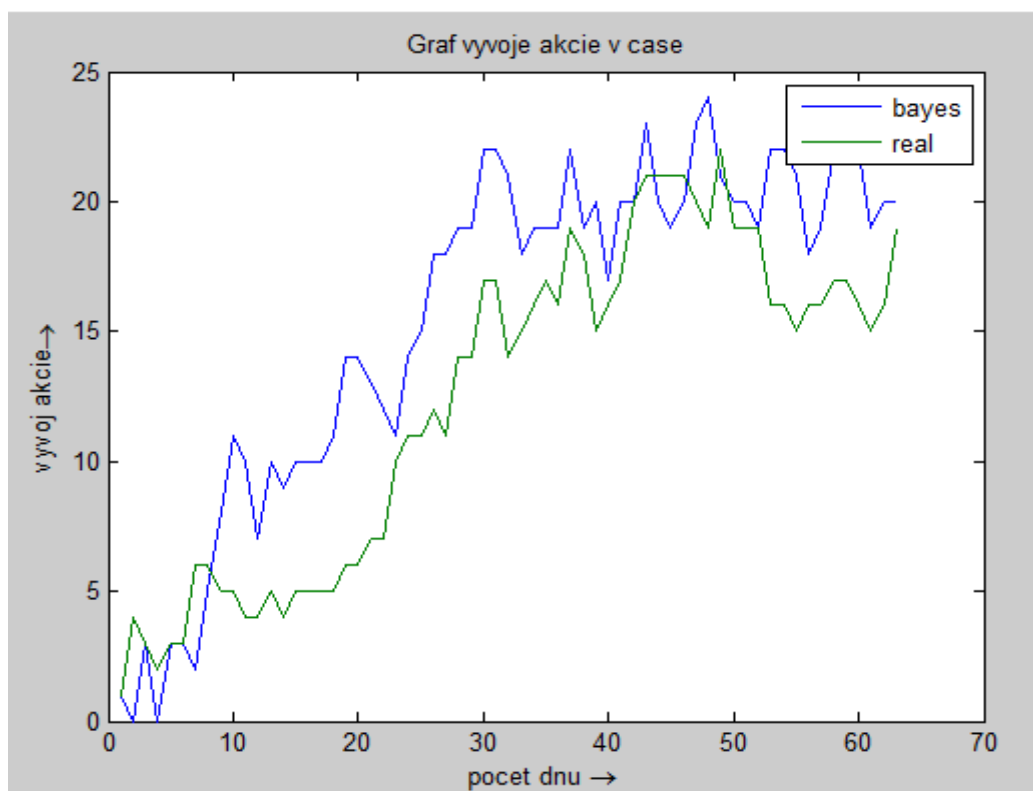
3.6 TEST 2 – 3x9

| | |
|---------------|---------|
| Délka pokusu | 9 týdnů |
| Trénování | 3 týdny |
| Počet iterací | 315 |

Druhá sada testů je determinována těmito parametry: délka trénování 3 týdny, což znamená ze se vytvoří trénování množina o 21 prvcích (dnech). Následný proces učení a predikce trvá 9 týdnů. Vlastní odhad vývoje ceny akcie je tedy porovnáván z hodnot naměřených v 63 dnech.

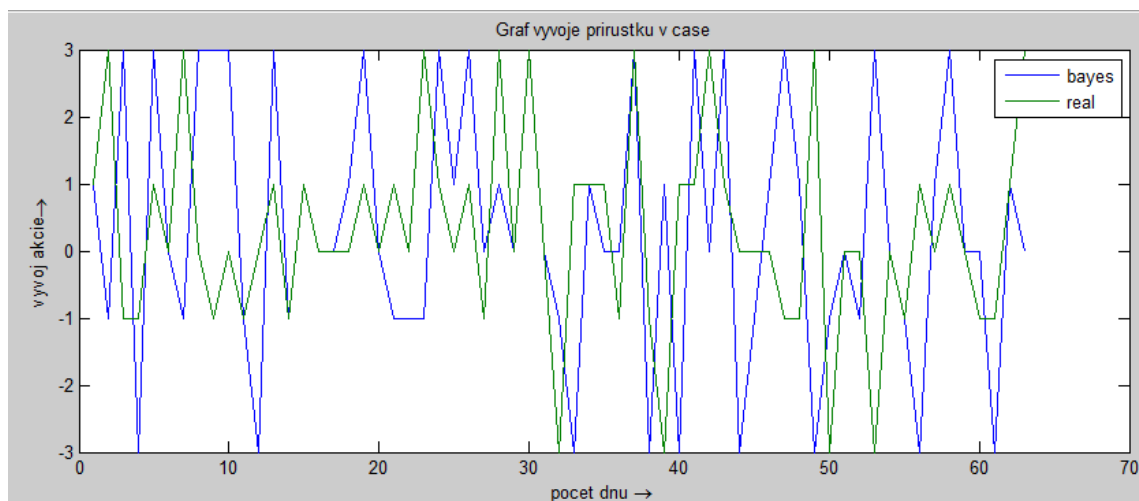
3.6.1 TEST 2.1

Při tak velkém počtu dnů a iterací je vzhledem k nedeterminovaným datům problém dosáhnout kvalitních výsledků. Následující příklad je ukázkou vhodné klasifikace, kdy byly důležité vztahy vyhodnoceny správně a predikce je proto úspěšná.



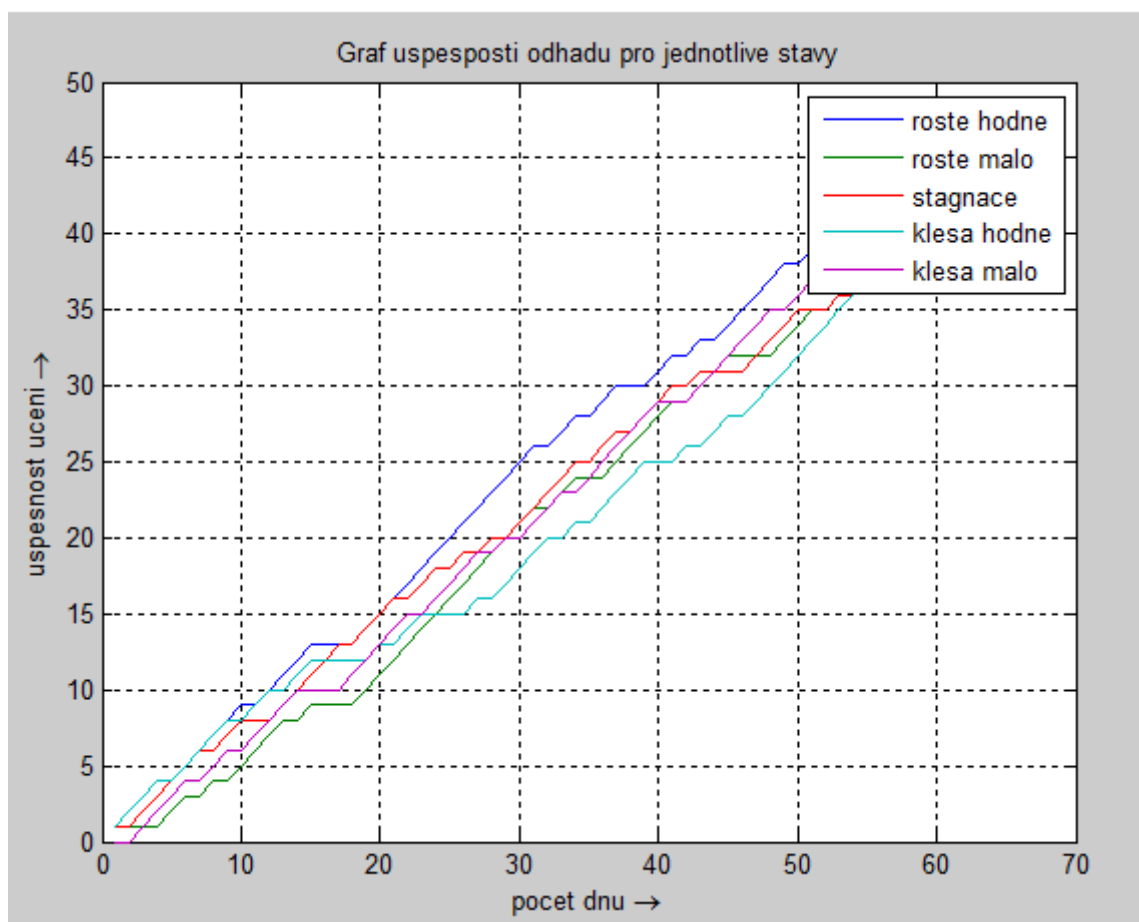
Obrázek 17: TEST 2.1 - vývoj akcie v čase - 3x9

Graf přírůstků reálných (obr. 18) a odhadovaných je již při takovém množství odhadovaných dní poněkud nepřehledný, nicméně lze pozorovat jistou úspěšnost s občasnými výkyvy v predikci.



Obrázek 18: TEST 2.1 - vývoj přírůstků v čase - 3x9

Procentuelní úspěšnost výstupních stavů (obr. 19) se pohybuje mezi 50 a 75 procenty.



Obrázek 19: TEST 2.1 – ukazatel úspěšnosti pro jednotlivé výstupy 3x9

3.7 Zhodnocení obecné části

Podle předešlých výsledů lze tvrdit, že pravděpodobností predikce pomocí Naivního Bayesova Klasifikátoru má reálnou šanci na úspěch. Výstup obecné části byl zatížen fiktivními daty, která byla vygenerována náhodně a nepopisovala tedy skutečnost jako takovou. Datový soubor však obsahoval normované vstupy a výstupy, které jsou pro klasifikaci a vyhodnocení závislostí nezbytné. Na základě těchto klasifikací byly provedeny úspěšné pokusy, kde predikce dosahovala přijatelných výsledků. Nicméně všechny pokusy pochopitelně vydařené nebyly, neboť při volbě trénovací množiny docházelo také k vybrání takových vstupních sad parametrů, které neměly reálný vliv na výstup, jinými slovy, docházelo k jevu, že pro určitou vstupní kombinaci parametrů existují různé výstupy v různých časových úsecích a to se projeví chybnou predikcí, kdy má podle předešlých výpočtů větší váhu odhad výstupu, který v dané situaci nenastane.

3.8 NBK – použití pro reálná burzovní data

Vzhledem k uspokojivým výsledkům předchozí části mohu přistoupit k testování algoritmu na reálných datech existující akcie. V první řadě bude nutno zvolit akcii takovou, která se chová v delším období alespoň v rámci daných pravidel relativně předpovědatelně, a to zejména z důvodu omezeného počtu vstupních parametrů algoritmu. Dále je nutné zvolit takové indikátory technické analýzy, mezi kterými bude algoritmus schopen vyhodnotit třídy závislostí, na kterých může predikce vývoje stavět. Pokud zvolíme nevhodné indikátory, může dojít k tomu, že datový soubor nebude ani z části deterministický, což je základním předpokladem pro použití pravděpodobnosti pro predikci, ale bude stochastický, čímž veškeré vypočítané závislosti a pravděpodobnosti parametrů nebudou odpovídat skutečnosti, neboť v případě stochastického souboru ani žádné zřejmé závislosti neexistují.

3.8.1 Kandidát pro predikci

Jako testovanou subjekt jsem zvolil globální společnost IBM. Důvodem je stabilita a rostoucí trend společnosti, což je vhodné pro klasifikaci při malém počtu vstupních faktorů a slibuje tak možný úspěch odhadu. Dokonce by se dalo říct, že společnost IBM na celosvětové finanční krizi vydělala a upevnila svou pozici na trhu.

IBM začne v následujících letech hromadně skupovat zpět své akcie. Do jara příštího roku by to mělo být další 10 miliard dolarů (v září nakoupili za 2,3 miliardy), během pěti let chtějí od svých akcionářů nakoupit cenné papíry v hodnotě 50 miliard.

Samuel J. Palmisano už od svého nástupu do čela společnosti v roce 2003 (řídil ji už v roce 2002) podporuje tuto strategii a v posledních sedmi letech nakoupili zpět akcie v hodnotě 91 miliard dolarů. Ve druhém čtvrtletí také podnik vyměřil výši dividend, která činí 75 centů.

A proč IBM kupuje zpět své akcie? Většinou to bývá proto, aby společnost navýšila jejich tržní cenu. A také proto, že si to může dovolit. IBM má smělé plány, když v pěti letech chce utratit 50 miliard za akcie, 20 miliard má připravených na nákupy a ještě musí čtvrtletně vyplácet dividendy z 1,24 miliard akcií (asi 800 mil. USD).

V současnosti se navíc firmě daří a hodnota se dostává opět na historická maxima okolo 170 dolarů za akcii. Tržní hodnota je v tuto chvíli 207 miliard USD, do pěti let by měla firma vlastnit dalších téměř 30 % svých akcií.

3.8.2 Data

Pro získání burzovních historických dat jsem použil funkci GoogleFinance v prostředí online dokumentů GoogleDocs. Postup pro naplnění tabulkového kalkulátoru byl jednoduchý. V první buňce listu jsem zapsal funkci pro zobrazení burzovních dat s následujícími parametry:

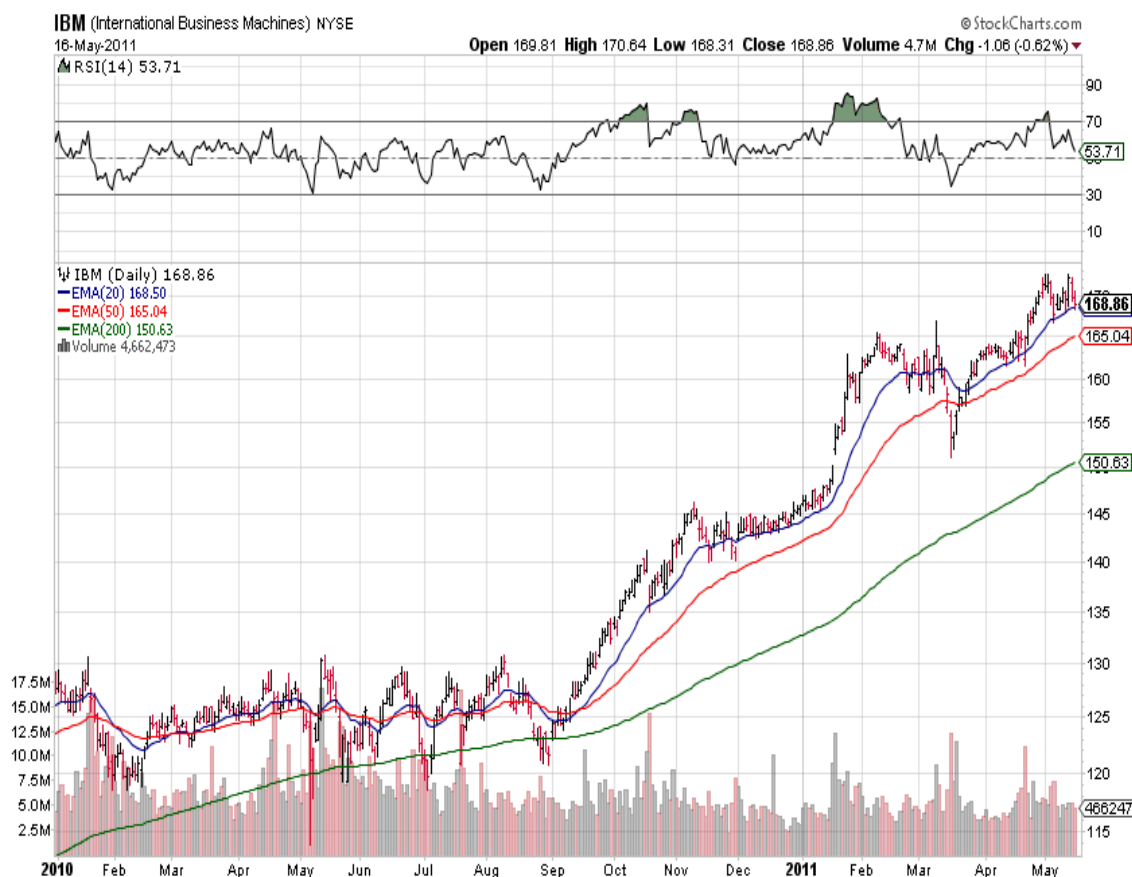
```
=GoogleFinance("IBM","all","2010-01-01","2011-05-15")
```

Kde prvním parametrem je symbol titulu (IBM), druhým je požadovaná hodnota – open, high, low, close nebo volume (výraz „all“ vypíše všech 5) a třetí a čtvrtý parametr označuje časový interval od-do. Výstupem funkce je tedy soubor 5 základních údajů o akciích v období 350 dní a je zobrazen na obr. 20.

| File Edit View Insert Format Data Tools Help | | | | | | |
|--|--------------------|--------|--------|---------|----------|----------|
| <div> </div> | | | | | | |
| Formula: =GoogleFinance("IBM","all","2010-01-01","2011-05-15") | | | | | | |
| | A | B | C | D | E | F |
| 1 | Date | Open | Close | High | Low | Volume |
| 68 | 4/5/2010 16:00:00 | 128.38 | 129.35 | 129.8 | 128.14 | 4118884 |
| 69 | 4/6/2010 16:00:00 | 128.68 | 128.93 | 129.3 | 128.05 | 3926409 |
| 70 | 4/7/2010 16:00:00 | 128.53 | 128.48 | 129.27 | 128.01 | 5157358 |
| 71 | 4/8/2010 16:00:00 | 128.04 | 127.61 | 128.23 | 127.2 | 6006833 |
| 72 | 4/9/2010 16:00:00 | 127.88 | 128.76 | 128.87 | 127.12 | 5186516 |
| 73 | 4/12/2010 16:00:00 | 128.57 | 128.36 | 128.956 | 128.24 | 3994168 |
| 74 | 4/13/2010 16:00:00 | 128.26 | 129.03 | 129.435 | 127.84 | 6821802 |
| 75 | 4/14/2010 16:00:00 | 129.73 | 131.25 | 131.42 | 129.46 | 8545857 |
| 76 | 4/15/2010 16:00:00 | 130.53 | 130.89 | 131.14 | 130.1902 | 6425399 |
| 77 | 4/16/2010 16:00:00 | 130.68 | 130.63 | 132.17 | 130.25 | 9549798 |
| 78 | 4/19/2010 16:00:00 | 130.38 | 132.23 | 132.28 | 130.38 | 11353679 |
| 79 | 4/20/2010 16:00:00 | 129.2 | 129.69 | 130.33 | 128.26 | 15218573 |
| 80 | 4/21/2010 16:00:00 | 129.87 | 128.99 | 130.27 | 128.5 | 7560015 |
| 81 | 4/22/2010 16:00:00 | 128.64 | 129.13 | 129.36 | 127.77 | 6018690 |
| 82 | 4/23/2010 16:00:00 | 129.08 | 129.99 | 130.1 | 128.71 | 6197382 |
| 83 | 4/26/2010 16:00:00 | 129.76 | 130.73 | 131.04 | 129.54 | 5285484 |
| 84 | 4/27/2010 16:00:00 | 129.9 | 128.82 | 132 | 128.71 | 10917510 |
| 85 | 4/28/2010 16:00:00 | 129.4 | 130.1 | 130.47 | 129.03 | 7123602 |
| 86 | 4/29/2010 16:00:00 | 130.55 | 130.46 | 131.21 | 130.15 | 5786888 |
| 87 | 4/30/2010 16:00:00 | 130.43 | 129 | 130.636 | 128.84 | 6266689 |
| 88 | 5/3/2010 16:00:00 | 129.39 | 129.6 | 130.14 | 128.8 | 4992019 |
| 89 | 5/4/2010 16:00:00 | 128.89 | 128.12 | 128.93 | 126.5754 | 8285289 |
| 90 | 5/5/2010 16:00:00 | 127.12 | 127.46 | 128.23 | 126.87 | 6072834 |

Obrázek 20: Data pořízená funkcí GoogleFinance

Z takto získaných dat je dále potřeba získat indikátory EMA20 a EMA50. Od použití indikátoru EMA 200 jsem nakonec ustoupil, protože je příliš „pomalý“ a v celém sledovaném období zasáhne významně pouze několikrát, jak je vidět z následujícího grafu, kde modrá čára představuje indikátor EMA20, červená EMA50 a zelená EMA200. Proto jeho použití pro predikci ve zvoleném období nemá smysl, jelikož by se normovaná hodnota tohoto indikátoru nemusela v predikci vůbec projevit.



Obrázek 21: Vývoj akcií IBM ve sledovaném období

Vzorec pro výpočet ukazatelů EMA je popsán v kapitole 1.4.5. Pro ukazatel EMA20 je prvních 20 hodnot použit ukazatel SMA, pro ukazatel EMA50 je prvních 50 hodnot ukazatel EMA20. Následně je vypočten rozdíl mezi $EMA_{20,t}$ a $EMA_{20,t-1}$, stejně tak pro $EMA_{50,t}$ a $EMA_{50,t-1}$. Získal jsem tak rozdíl mezi současnou a minulou hodnotou ukazatele a jsem tedy schopen říci, zda-li indikátor roste či klesá.

Dalším ze vstupních parametrů je rozdíl P_t a P_{t-1} . Ten nám určuje aktuální cenový trend, tedy rostoucí, klesající nebo stagnace.

3.8.3 Normalizace dat

Prvním vstupním parametrem je den v týdnu. Téměř všechny obchody probíhají v období od pondělí do pátku. V sobotu lze sice také obchodovat, nicméně toto obchodování je značně omezeno. Proto parametr den v týdnu nabývá hodnot pouze od 1 do 5. Den v týdnu sice není technickým parametrem, nicméně jsem ho do vstupů

zahrnul. V další optimalizaci algoritmu ho doporučuji nahradit dalším technickým parametrem, protože efekt pátečního zvýšeného obchodování nebo například pondělní nervozity postupem let mizí.

Hodnoty vypočtených diferencí indikátorů EMA20 a EMA50, společně s diferencí ceny je potřeba převést do formy, se kterou může algoritmus pracovat, tzn. převést je do celočíselné formy.

Pro určení stagnace bylo potřeba použití intervalu, protože rozdíl mezi současnou a předchozí hodnotou není 0. V závislosti na maximálním/minimálním přírůstku/poklesu ceny, které činily ve sledovaném období +5,79 a -6,02 jsem zvolil interval pro stagnaci -0,3 – 0,3. Pro indikátory EMA byl interval stagnace -0,05 – 0,05.

Indikátory EMA20 a EMA50 jsou převedeny následovně:

- Klesá – pokud je rozdíl mezi dnešní a včerejší hodnotou menší než 0,05, pak normovaný vstup je -1.
- Stagnuje – pokud je rozdíl mezi dnešní a včerejší hodnotou v intervalu $-0,05 \div 0,05$, pak normovaný vstup je 0.
- Roste – pokud je rozdíl mezi dnešní a včerejší hodnotou větší než 0,05, pak normovaný vstup je 1.

Parametr rozdílu ceny P_t a P_{t-1} nemá pouze 3 možné hodnoty výstupu, ale 5:

- Klesá hodně – pokud je rozdíl mezi dnešní a včerejší hodnotou menší jak -2, pak normovaný vstup je -3.
- Klesá málo – pokud je rozdíl mezi dnešní a včerejší hodnotou menší jak -0,3, pak normovaný vstup je -1.
- Stagnuje – pokud je rozdíl mezi dnešní a včerejší hodnotou v intervalu -0,3 – 0,3, pak normovaný vstup je 0.
- Roste málo – pokud je rozdíl mezi dnešní a včerejší hodnotou větší jak 0,3 a zároveň menší jak 2, pak normovaný výstup je 1.

- Roste hodně – pokud je rozdíl mezi dnešní a včerejší hodnotou větší jak 2, pak normovaný vstup je 3.

Část datového souboru je vidět na následujícím obrázku.

| M | N | O | P | Q | R | S | T | U |
|-------------------|-------------------|-------------------|-------------------|----------|----------|----------|----------|----------|
| normovaný vstup 1 | normovaný vstup 2 | normovaný vstup 3 | normovaný vstup 4 | vystup 1 | vystup 2 | vystup 3 | vystup 4 | vystup 5 |
| 1 | 1 | 1 | -1 | 0 | 0 | 0 | 1 | 0 |
| 2 | -1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | -3 | 0 | 1 | 0 | 0 | 0 |
| 4 | 1 | 1 | -1 | 0 | 1 | 0 | 0 | 0 |
| 5 | 1 | 1 | -1 | 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | -1 | 0 | 0 | 1 | 0 | 0 |
| 2 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| 4 | -1 | -1 | 3 | 0 | 0 | 0 | 1 | 0 |
| 5 | -1 | -1 | 1 | 0 | 0 | 1 | 0 | 0 |

Obrázek 22: datový soubor normovaných dat

3.9 Testování na reálných datech

Kompletní datový soubor s normovanými vstupy obsahuje 331 dní, tzn. 2979 hodnot vstupních a výstupních údajů. Tak velký vzorek dat by měl pro testování predikce poskytnout dostatek vstupních kombinací pro vyhodnocení závislosti pro predikci. Testování probíhá obdobě jako s fiktivními daty, nicméně proces trénování i učení je delší.

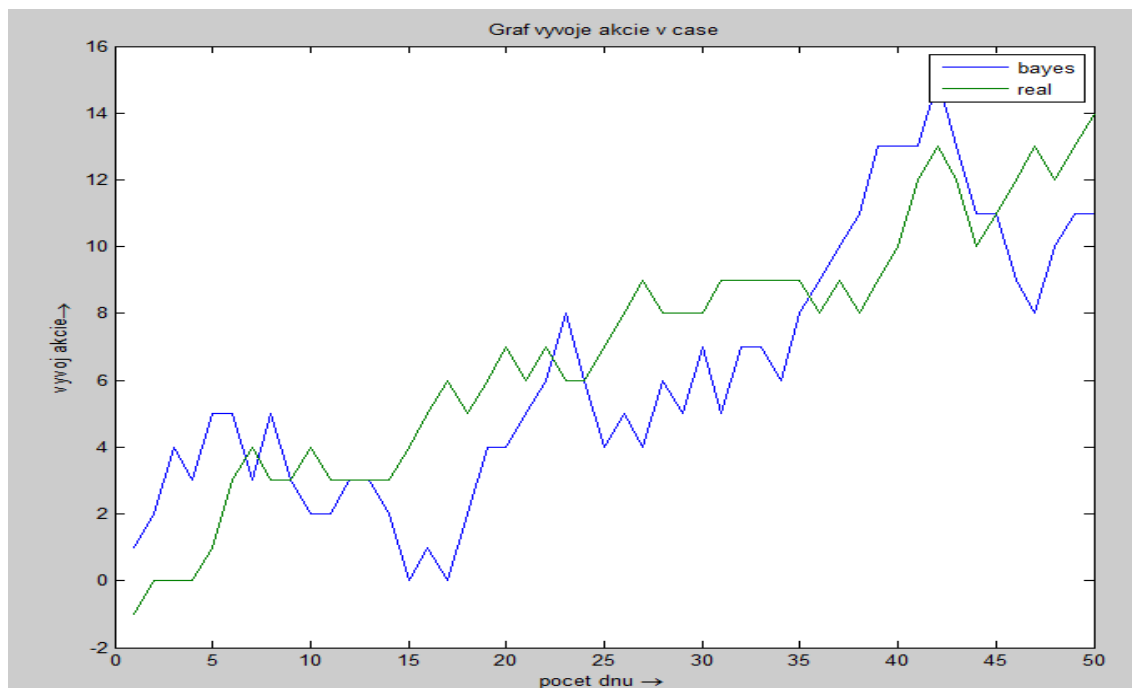
3.10 TEST 3 – 10x10

| | |
|---------------|----------|
| Délka pokusu | 10 týdnů |
| Trénování | 10 týdnů |
| Počet iterací | 250 |

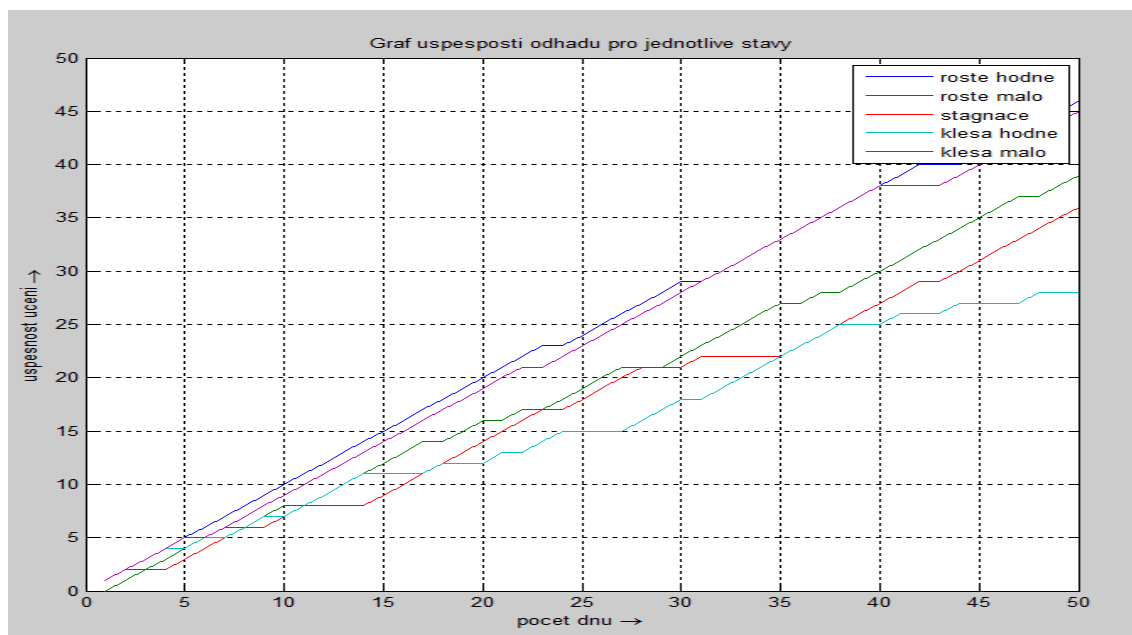
První sada testů je determinována těmito parametry: délka trénování 10 týdnů, což znamená že se vytvoří trénovací množina o 50 prvcích (dnech). Následný proces učení a predikce trvá dalších 10 týdnů. Vlastní odhad vývoje ceny akcie je tedy porovnáván z hodnot naměřených v 50 dnech.

3.10.1 TEST 3.1

Výsledek prvního zobrazeného pokusu je na obrázku 23. Jak je z grafu vidět, predikce s menšími či většími výkyvy osciluje kolem skutečného vývoje akcií IBM. Odhad začíná mírně nad skutečnou hodnotou v počátku predikce a končí mírně pod ní. Nicméně nelze řadit mezi ideální výsledky, vzhledem k celkovému průběhu.



Obrázek 23: TEST 3.1 - graf vývoje akcie 10x10



Obrázek 24: TEST 3.1 - Ukazatele úspěšnosti pro jednotlivé výstupy 10x10

Jelikož úspěšných výsledků nebylo mnoho a algoritmus s těmito parametry nedosahoval valných úspěchů, rozhodl jsem se neplýtvat místem a provedl jsem změnu běhových parametrů. Stejná velikost trénovací množiny a délky pokusu je podle dosahovaných výsledků nedostatečná a algoritmus nestihne v tak krátkém běhu vytvořit dostatek závislostí a nedochází tedy ke korelaci vstupu a výstupu. Další sada testů bude tedy s pokusem delším než je velikost trénovací množiny.

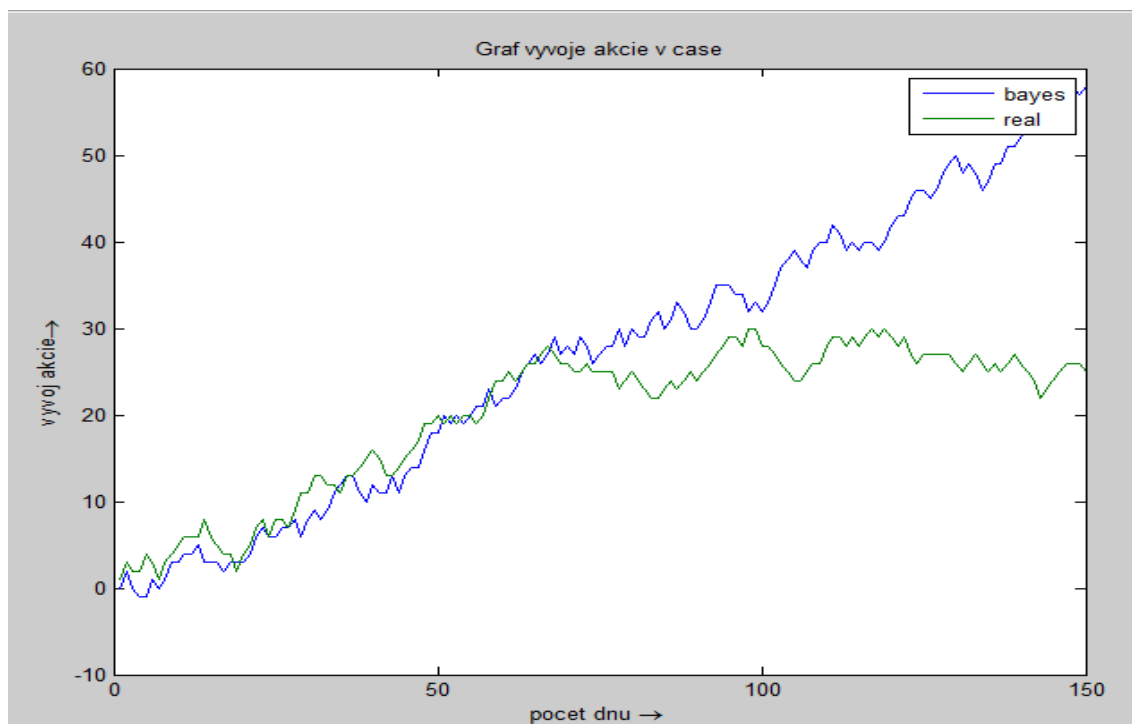
3.11 TEST 4 – 5x30

| | |
|---------------|----------|
| Délka pokusu | 30 týdnů |
| Trénování | 5 týdnů |
| Počet iterací | 750 |

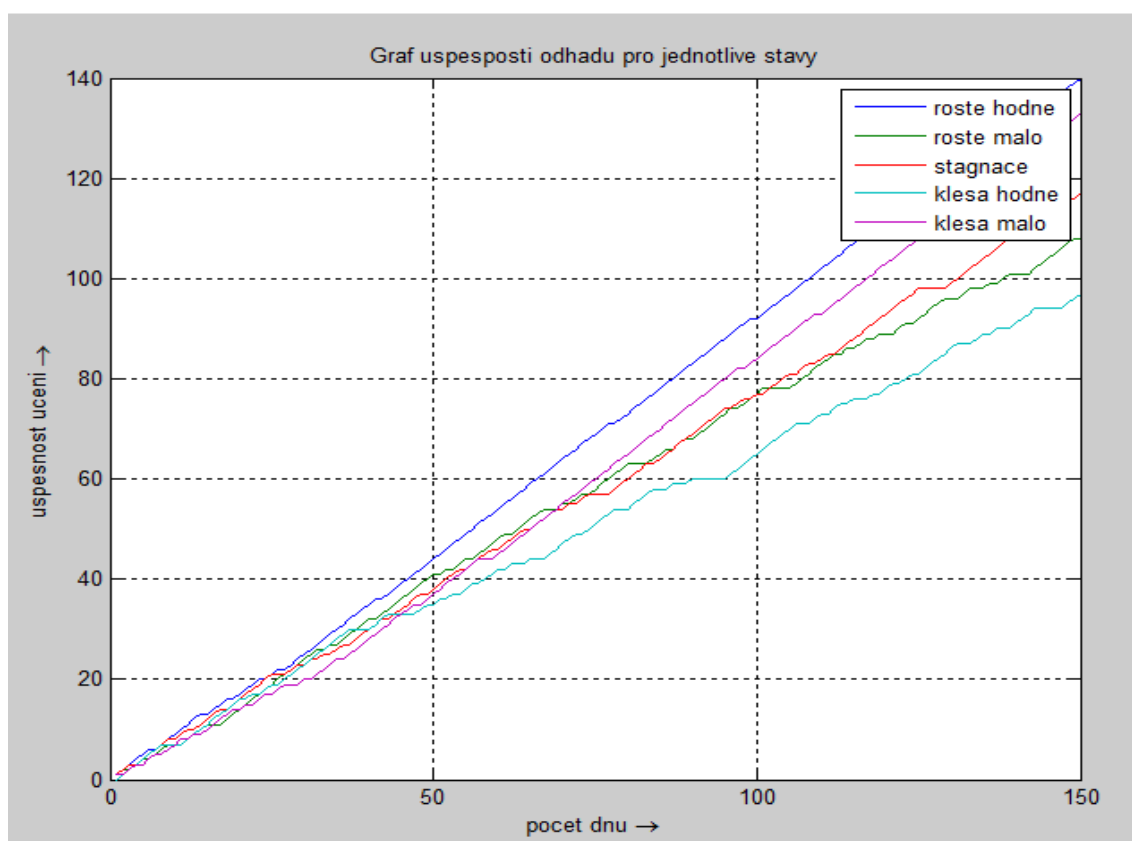
Druhá sada testů je determinována těmito parametry: délka trénování 5 týdnů, což znamená, že se vytvoří trénovací množina o 25 prvcích (dnech). Následný proces učení a predikce trvá dalších 30 týdnů. Vlastní odhad vývoje ceny akcie je tedy porovnáván z hodnot naměřených v 150 dnech.

3.11.1 TEST 4.1 a 4.2

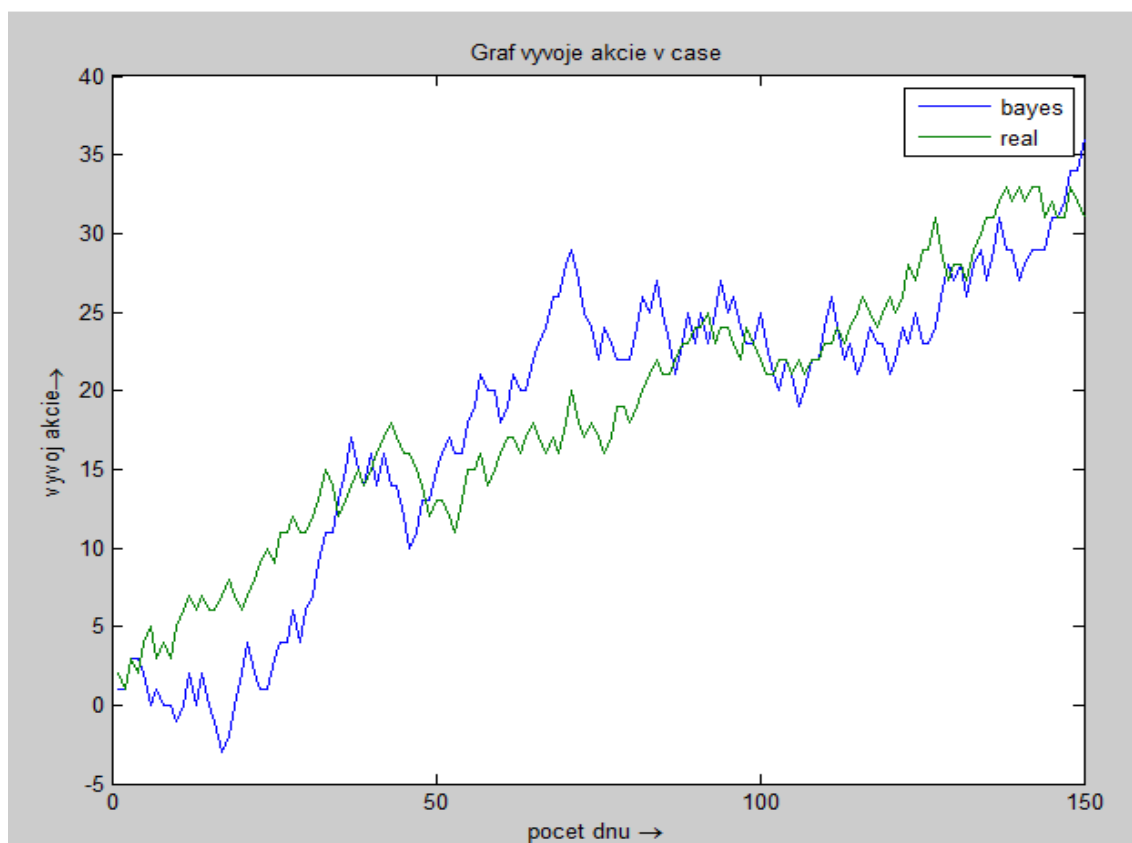
Tato volba parametrů délky trénování a celkové délky pokusu vedla k zajímavým výsledkům. Většina odhadů proběhla s dobrou přesností, jak je vidět na následujících grafech. Přes drobné nedostatky, jak je tomu vidět například na obrázku 25 dosahovala úspěšnost odhadu 60 až 80 procent v 1/3 průběhu. Ve dvou třetinách se horní hranice úspěšnosti ukazatelů stále blížila k 75 procentům a u některých ukazatelů tato úspěšnost vydržela až do konce testu. Výkyvy v predikci jsou však samozřejmostí, vzhledem k relativně malé množině vstupních parametrů determinujících výstup.



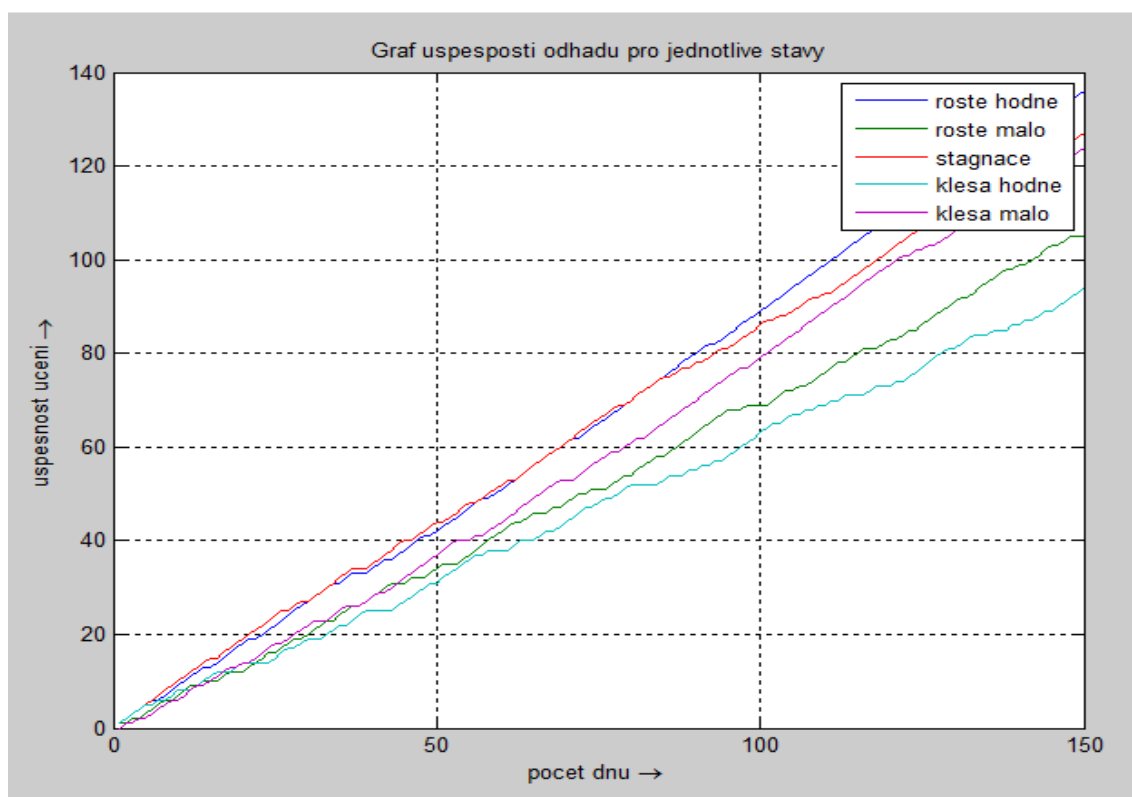
Obrázek 25: TEST 4.1 - graf vývoje akcie 5x30



Obrázek 26: TEST 4.1 - Ukazatele úspěšnosti pro jednotlivé výstupy 5x30



Obrázek 27: TEST 4.2 - graf vývoje akcie 5x30



Obrázek 28: TEST 4.2 - Ukazatele úspěšnosti pro jednotlivé výstupy 5x30

3.11.2 TEST 5 – 5x60

Poslední sada testů je prováděna s velmi dlouhým trváním pokusu se zachováním trénovací množiny. Pokus je tedy v délce 60 týdnů, což je 300 dní. Vypustím pro tuto sadu grafy úspěšnosti ukazatelů a ponechám pouze vlastní predikci vývoje.

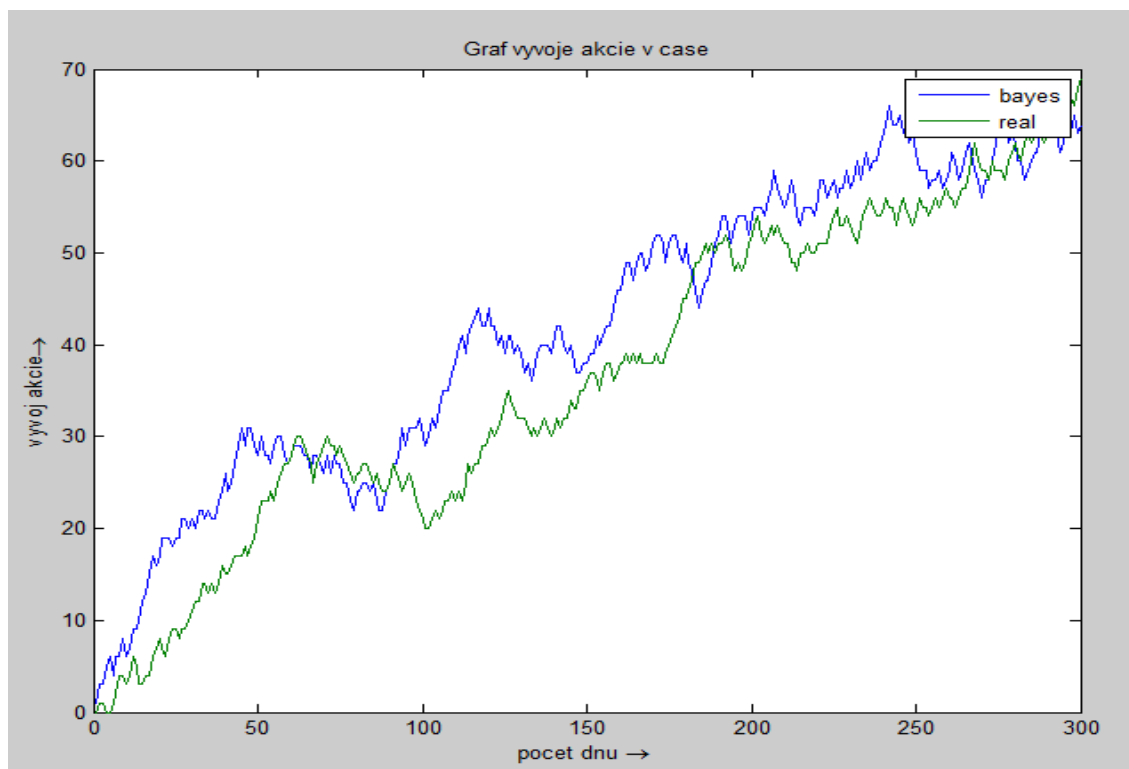
Nutno podotknout, že pro takto dlouhý pokus již výsledky nebyly natolik přesvědčující jako v předchozích testech, nicméně to opět může být způsobeno slabou závislostí vstupů.

Trénovací množina je, jak jsem již zmínil, vybírána náhodně. To by mohlo mít vliv na výstup v tom smyslu, že data vybraná do trénovací množiny mohou být v úplně jiné části datového souboru, řekněme rok před daty, která se předhazují algoritmu pro predikci. V takovém případě může dojít k situaci, že naučené závislosti v novém období nemají již vliv na výstup. Potom modelu samozřejmě chvíli trvá, než upraví váhy pravděpodobností tak, aby se zvyšovala úspěšnost odhadu. V následujících grafech je tento jev viditelný.

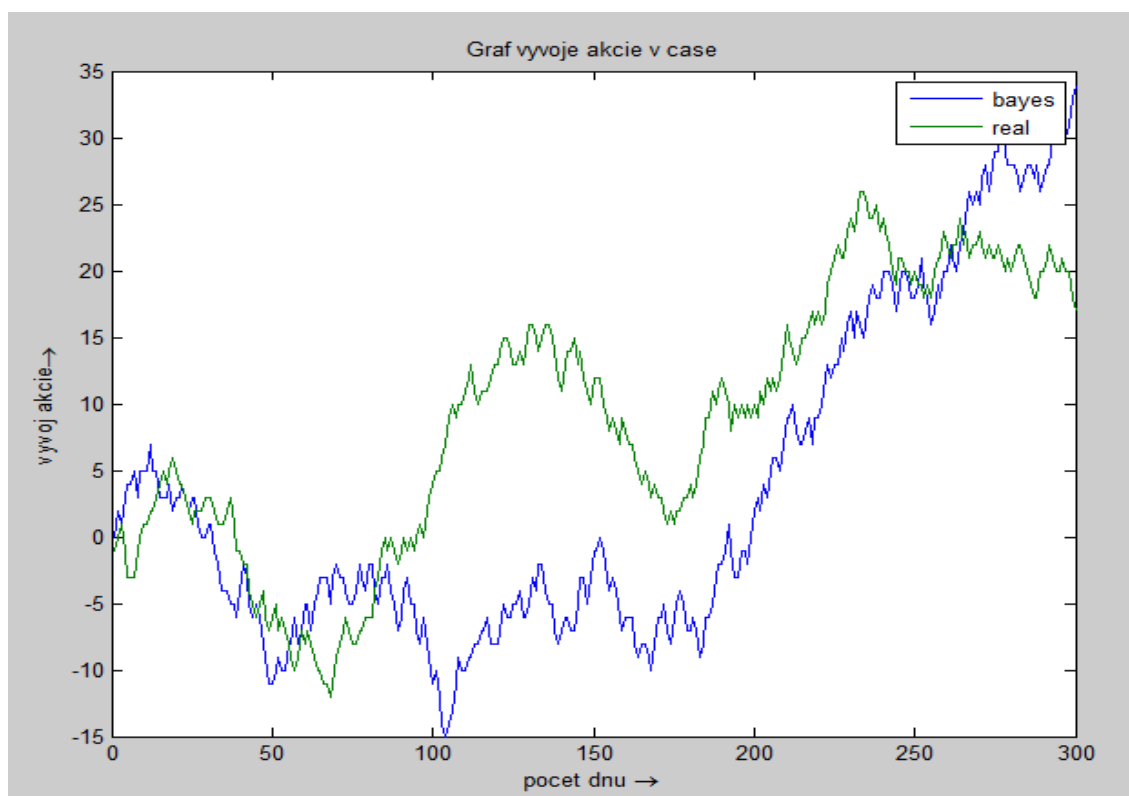
Na obrázku 29 je vidět velmi zajímavý jev, a to předstih před skutečností, což by v realitě byla naopak vlastnost velmi žádoucí. Predikce se však po třech velmi zajímavých vlnách ustálí a postupně se blíží skutečné hodnotě.

Obrázek 30 je druhým zástupcem této sady testů. Predikce začíná úspěšně, pak následuje chybné vyhodnocení ukazatelů, které může být způsobeno trénovacími daty, jak je zmíněno výše. Po určité době se však opět odhad přiblíží ke skutečnosti.

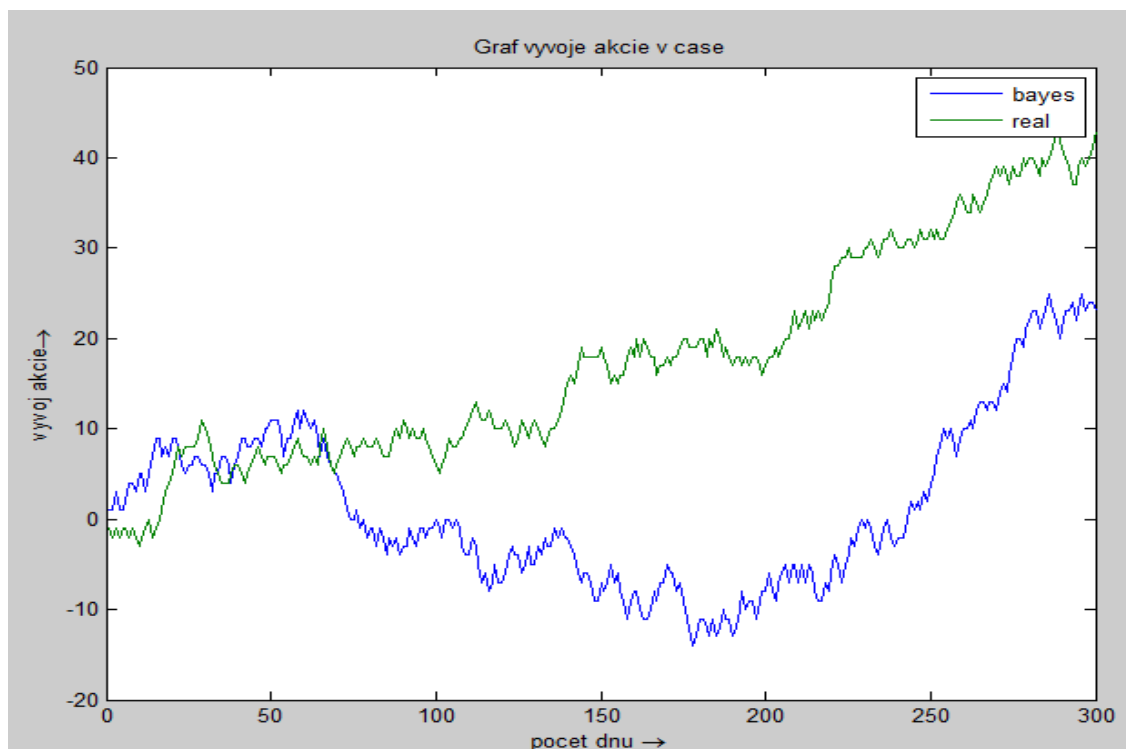
Posledním testem na reálných datech je graf na obrázku 31. Zde je patrné špatné vyhodnocení vstupů a s tím spojený propad mezi 60. a 70. dnem. V dalším období se však predikce uklidní a v podstatě kopíruje skutečný vývoj, jen je poznamenána prvotním pochybením.



Obrázek 29: TEST 5.1 - Ukazatele graf vývoje akcie 5x60



Obrázek 30: TEST 5.2 - Ukazatele graf vývoje akcie 5x60



Obrázek 31: TEST 5.3 - Ukazatele graf vývoje akcie 5x60

3.12 Praktické použití pro investice

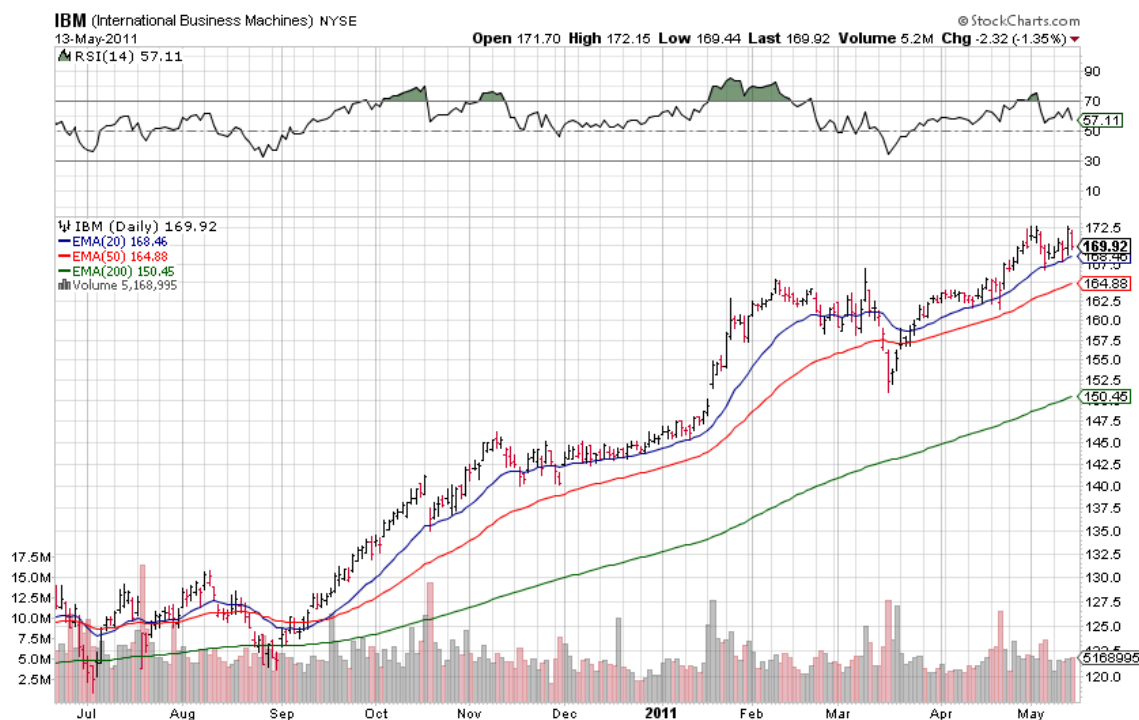
V této kapitole se pokusím přiblížit použití programu v modelové situaci. Půjde o použití na skutečných datech IBM, nákup v určité fázi predikce a výpočet výnosu/ztráty, které bychom hypoteticky dosáhli, pokud by jsme se řídili odhadem, který nám předloží Naivní Bayesův Klasifikátor. Samozřejmě je potřeba vybrat takový odhad, který za dobu učení projeví určitou úspěšnost, a je tedy možné v době nákupu předpokládat úspěšnost odhadu i pro další období.

Základem je volba sledovaného období. Zvolil jsem interval od 22. 6. 2010 do 13. 5. 2011, kde akcie IBM mají rostoucí tendenci a pomalu se smazává zatíženost indikátorů EMA z předchozích období. Nastavení algoritmu je pro pokus následující:

Velikost trénovací množiny: 5 týdnů

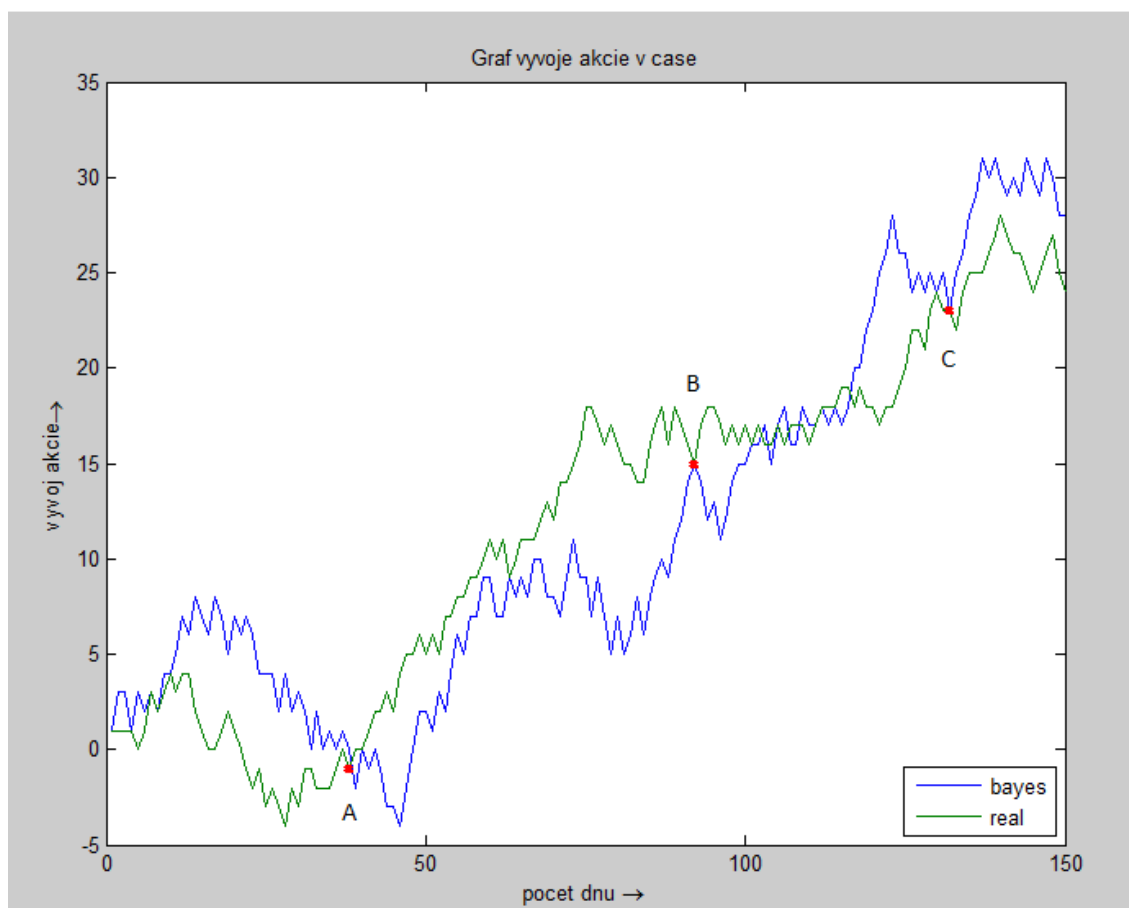
Délka pokusu: 30 týdnů

Start predikce: 22. 6. 2010



Obrázek 32: Vývoj IBM ve zvoleném období 22/6/2010 – 13/5/2011

Jak je vidět z následujícího grafu predikce, algoritmus kopíruje cenový trend s mírným zpožděním. Je to způsobeno nepřesnou predikcí v počátku. Nicméně odhad protne skutečný vývoj v bodě A, což z předchozího grafu odpovídá odrazu ceny akcie od velmi pomalého indikátoru EMA 200, tedy exponenciálního klouzavého průměru za posledních 200 dní. Tento indikátor je obchodníky interpretován jako jakási hranice, pod kterou se akcie bez zjevných ekonomických důvodů neudrží. Algoritmus proto po doběhnutí zatížení z nepřesného odhadu doběhne do odhadovaného minima, které odpovídá minimu akcie v předchozích dnech (dáno vyhodnocením EMA 20 a EMA 50). Bod protnutí ceny odhadnuté a skutečné je možno vzít jako signál k nákupu, k čemuž přispívá skutečnost krátkodobého pádu pod EMA 200.



Obrázek 33: Predikce a skutečnost normovaných dat v období

Obtížnější je hledat signál k prodeji. K protnutí ceny predikované s cenou skutečnou ze zdola nahoru je až v období stagnace. Zde je třeba vyhodnotit další indikátory, jako kupříkladu účetní závěrky, výroční zprávy nebo dění kolem společnosti. Program nicméně po období stagnace předpovídá na základě změn indikátorů EMA 20 a EMA 50 růst, což se v zápětí skutečně děje. Doporučit proto vhodný okamžik k prodeji není v dané situaci snadné, ať již z důvodu predikce růstu, tak k celkovému trendu za období.

Můžeme proto pro výpočet potenciálního výnosu z investice použít pevně danou délku držení akcie, popřípadě předem stanovený výnos. Prodej předpokládám za cenu Close, nákup za cenu Open.

Pro pevně stanovenou délku držení akcie 30/60/90 dní jsou údaje následující:

- Na počátku období (22.6.2010) byly denní hodnoty akcie IBM tyto:

Open 130.37, Close 129.3, High 131.47, Low 129.07

- V době nákupu (13.8.2010), tedy v bodě protnutí ceny skutečné a predikce:

Open 127.96, Close 127.87, High 128.46, Low 127.33

| Délka držení | 30 | 60 | 90 |
|----------------------|-----------|-----------|------------|
| Datum nákupu | 13.8.2010 | 13.8.2010 | 13.8.2010 |
| Datum prodeje | 24.9.2010 | 5.11.2010 | 20.12.2010 |
| Nákup | 127,96 | 127,96 | 127,96 |
| Prodej | 134,11 | 146,92 | 144,51 |
| Zhodnocení | 4,8 % | 14,8 % | 12,9 % |

Závěr

Podle provedených pokusů a jejich výsledků lze tvrdit, že užití pravděpodobnostní metody pro predikci vývoje akcií, potažmo i pro predikci jakýchkoli jiných jevů závislých na determinismu, může dosahovat zajímavých výsledků za předpokladu správné identifikace závislostí výstupu na vstupních podmínkách. Naivní Bayesovský klasifikátor použitý v této diplomové práci prokázal, že by mohl být metodou, která poslouží stejně dobře, jako notoricky známé užití neuronových sítí. Ba co víc, NBK by nemusel být ani tak konkurencí pro neuronové sítě, ale naopak, spojením těchto dvou rozdílných disciplín by mohl vzniknout velmi silný nástroj predikce. Proto také v dnešní době roste zájem o BK a jsou intenzivně zkoumány pro svou schopnost klasifikace při relativně nízké výpočetní náročnosti. Umožňuje to tak efektivně identifikovat vzájemné vztahy mezi objekty, které jsou životně důležité pro úspěšné předpovídání jevů.

Nástroj vytvořený v této diplomové práci může sloužit kupříkladu jako jeden z mnoha analytických nástrojů pro investiční rozhodování. Před použitím je však třeba zvážit, které vstupní parametry budou pro dané využití nejvhodnější. Vytvořený algoritmus jako takový samozřejmě není všelékem. Pro specifická použití proto vyžaduje specifická nastavení. Životně důležitou fází je právě ona volba vstupních parametrů a jejich kvalitní normalizace. Je totiž důležité najít takové vstupní parametry, na základě kterých algoritmus vyhodnotí správné vztahy mezi vstupem a výstupem.

Problémem u akciových trhů, potažmo všech pokusů o jejich predikci, je jejich stochastická složka. Tzn. změny vyvolané na trhu, které buď nemají logickou ekonomickou příčinu, nebo se jedná o nepředvídatelný jev. Příkladem budiž falšování ekonomických výsledků společnosti nebo spekulativní nákupy/prodeje v případě zneužití informací v obchodním styku. S těmito a jim podobnými faktory nemůže pochopitelně žádná predikce počítat.

Použití algoritmu bych doporučil jako nástroj pro podporu rozhodování například ve společnosti, která uvažuje využití části svého kapitálu v oblasti spekulativního obchodování. Přestože velmi závisí na dobré identifikaci relevantních vstupů pro model, může společnost dosáhnout při dostatečném testování zajímavých výsledů. Po delším zkoumání trhu je samozřejmě možnost vybrat i větší množství vstupních faktorů, popřípadě zvětšit šířku kteréhokoli vstupu pro lepší pokrytí vznikajících situací.

Seznam použité literatury

- 1) AKCIOVÉ ANALÝZY [online].[cit. 2011-10-05] Dostupné z WWW:
http://www.vaulton.webpark.cz/vauteorie/_method/M32_akcie.html
- 2) An Intuitive Explanation of Bayesian Reasoning [online] Eliezer S. Yudkowsky [cit. 2011-6-5]. Dostupné z WWW: <<http://yudkowsky.net/rational/bayes>>
- 3) Bayes theorem [online].[cit. 2011-6-5]. Dostupné z WWW:
<http://en.wikipedia.org/wiki/Bayes_theorem>
- 4) BRADA, J. Technická analýza. Praha: VŠE, 2000. ISBN 80-245-0096-5.
- 5) DOMINGOS, P., PAZZANI, M. (1997), On the optimality of the simple Bayesian classifier under zero-one loss, Machine Learning, vol. 29, str. 103-130.
- 6) HONZÍK, P. Strojové učení. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2006. 85s.
- 7) Metoda nejmenších čtverců. [online].[cit. 2011-7-5] Dostupné z WWW:
<<http://user.mendelu.cz/marik/mat-web/mat-webse24.html>>
- 8) Naive bayes Classifier [online].[cit. 2011-27-4]. Dostupné z WWW:
<http://en.wikipedia.org/wiki/Naive_Bayes_classifier>
- 9) O Bayesovském učení. [online]. Josef Heřmann [cit. 2011-8-5]. Dostupné z WWW <http://www.odbornecasopisy.cz/index.php?id_document=28517>
- 10) PONSI, Ed. Forex Patterns & Probabilities. Hoboken, New Jersey : John Wiley & Sons, Inc., 2007. ISBN 987-0-470-09729-8. S. 33-130.
- 11) Psychologická analýza [online][cit. 2011-10-5]. Dostupné z WWW
< http://cs.wikipedia.org/wiki/Psychologická_analýza>
- 12) REJNUŠ, O: Peněžní ekonomie – Finanční trhy. Nakladatelství Cerm, Brno, 2008. Vydání čtvrté, 352 stran. ISBN 9788021437036
- 13) ROSE, P. S.: Peněžní a kapitálové trhy: Finanční systém ve stále globálnější ekonomice. Nakladatelství Victoria Publishing, Praha, 1994, Vydání první, 1014 stran. ISBN 808560552X

- 14) VESELÁ, J.: Investování na kapitálových trzích. Nakladatelství ASPI a.s., Praha, 2007. Vydání první, 704 stran. ISBN 9788073572976, str. 22-23

Seznam zkratek

| | |
|-----|--------------------------------|
| MA | Moving Average |
| SMA | Simple Moving Average |
| EMA | Exponential Moving Average |
| NBK | Naivní Bayesův Klasifikátor |
| BBK | Brutální Bayesův Klasifikátor |
| OBK | Optimální Bayesův Klasifikátor |
| SU | Strojové učení |
| ML | Machine Learning |
| UI | Umělá inteligence |
| AI | Artificial Intelligence |
| MSE | Minimum Square Error |
| UA | Učící algoritmus |
| LF | Loss Function |

Seznam obrázků

| | |
|--|----|
| Obrázek 1: Rozdělení finančního trhu | 15 |
| Obrázek 2: Zařazení oboru Strojové učení dle ing. P. Honzíka, Ph.D.(6)..... | 25 |
| Obrázek 3: Model učení s učitelem (6)..... | 28 |
| Obrázek 4: Typy veličin (6)..... | 29 |
| Obrázek 5: Metoda nejmenších čtverců (7). | 31 |
| Obrázek 6: Datový soubor pro NBK | 46 |
| Obrázek 7: Vývojový diagram NBK | 47 |
| Obrázek 8: TEST 1.1 - vývoj akcie v čase - 3x5 | 58 |
| Obrázek 9: TEST 1.1 – vývoj přírůstků v čase – 3x5..... | 59 |
| Obrázek 10: TEST 1.1 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5 | 60 |
| Obrázek 11: TEST 1.2 – vývoj akcie v čase – 3x5..... | 61 |
| Obrázek 12: TEST 1.2 - vývoj přírůstků v čase - 3x5 | 61 |
| Obrázek 13: TEST 1.2 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5 | 62 |
| Obrázek 14: TEST 1.3 - vývoj akcie v čase - 3x5 | 63 |
| Obrázek 15: TEST 1.3 - vývoj přírůstků v čase - 3x5 | 64 |
| Obrázek 16: TEST 1.3 – ukazatel úspěšnosti pro jednotlivé výstupy 3x5 | 65 |
| Obrázek 17: TEST 2.1 - vývoj akcie v čase - 3x9 | 66 |
| Obrázek 18: TEST 2.1 - vývoj přírůstků v čase - 3x9 | 67 |
| Obrázek 19: TEST 2.1 – ukazatel úspěšnosti pro jednotlivé výstupy 3x9 | 67 |
| Obrázek 20: Data pořízená funkcí GoogleFinance | 70 |
| Obrázek 21: Vývoj akcií IBM ve sledovaném období | 71 |
| Obrázek 22: datový soubor normovaných dat | 73 |
| Obrázek 23: TEST 3.1 - graf vývoje akcie 10x10 | 74 |
| Obrázek 24: TEST 3.1 - Ukazatele úspěšnosti pro jednotlivé výstupy 10x10 | 74 |
| Obrázek 25: TEST 4.1 - graf vývoje akcie 5x30 | 76 |
| Obrázek 26: TEST 4.1 - Ukazatele úspěšnosti pro jednotlivé výstupy 5x30 | 76 |
| Obrázek 27: TEST 4.2 - graf vývoje akcie 5x30 | 77 |
| Obrázek 28: TEST 4.2 - Ukazatele úspěšnosti pro jednotlivé výstupy 5x30 | 77 |
| Obrázek 29: TEST 5.1 - Ukazatele graf vývoje akcie 5x60 | 79 |
| Obrázek 30: TEST 5.2 - Ukazatele graf vývoje akcie 5x60 | 79 |
| Obrázek 31: TEST 5.3 - Ukazatele graf vývoje akcie 5x60 | 80 |
| Obrázek 32: Vývoj IBM ve zvoleném období 22/6/2010 – 13/5/2011 | 81 |
| Obrázek 33: Predikce a skutečnost normovaných dat v období | 82 |